(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 March 2002 (07.03.2002)

PCT

(10) International Publication Number
WO 02/18420 A2

(51) International Patent Classification[7]: C07K 14/00

(21) International Application Number: PCT/EP01/09488

(22) International Filing Date: 17 August 2001 (17.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00118634.5 28 August 2000 (28.08.2000) EP

(71) Applicant *(for all designated States except US)*: **LION BIOSCIENCE AG** [DE/DE]; Im Neuenheimer Feld 515, 69120 Heidelberg (DE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ALBERS, Michael** [DE/DE]; Obere Schulstrasse 3, 69221 Dossenheim (DE). **ELLWANGER, Silvia** [DE/DE]; Theodor-Heuss-Strasse 4, 69181 Leimen-St. Ilgen (DE). **KÖGL, Manfred** [DE/DE]; Hauptstr. 131/4, 69214 Eppelheim (DE). **LÖSER, Eva** [DE/DE]; Biethstrasse 12, 69121 Heidelberg (DE).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/18420 A2

(54) Title: NOVEL COFACTORS OF THE PREGNANE X RECEPTOR AND METHODS OF USE

(57) Abstract: The present invention relates to novel cofactors of the Pregnane X Receptor which we call CF1, CF2, CF3, CF4 and CF 44 the isolated nucleic acid sequences thereof and the isolated proteins thereof. The invention further relates to processes for isolating and/or producing the nucleic acids or the proteins as well as methods of use of these cofactors, such as inhibiting or activating the binding of the cofactors to PXR.

NOVEL COFACTORS OF THE PREGNANE X RECEPTOR AND METHODS OF USE

BACKGROUND OF THE INVENTION

Multicellular organisms are dependent on advanced mechanisms of information transfer between cells and body compartments. The information that is transmitted can be highly complex and can result in the alteration of genetic programs involved in cellular differentiation, proliferation, or reproduction. The signals, or hormones, are often simple molecules, such as peptides, fatty acid, or cholesterol derivatives.

Many of these signals produce their effects by ultimately changing the transcription of specific genes. One well-studied group of proteins that mediate a cell's response to a variety of signals is the family of transcription factors known as nuclear receptors, hereinafter referred to frequently as "NR". Members of this group include receptors for steroid hormones, vitamin D, ecdysone, cis and trans retinoic acid, thyroid hormone, bile acids, cholesterol-derivatives, fatty acids (and other peroxisomal proliferators), as well as so-called orphan receptors, proteins that are structurally similar to other members of this group, but for which no ligands are known (Escriva, H. et al., Ligand binding was acquired during evolution of nuclear receptors, PNAS, 94, 6803 – 6808, 1997). Orphan receptors may be indicative of unknown signaling pathways in the cell or may be nuclear receptors that function without ligand activation. The activation of transcription by some of these orphan receptors may occur in the absence of an exogenous ligand and/or through signal transduction pathways originating from the cell surface (Mangelsdorf, D. J. et al., The nuclear receptor superfamily: the second decade, Cell 83, 835-839, 1995).

In general, three functional domains have been defined in NRs. An amino terminal domain is believed to have some regulatory function. A DNA-binding domain hereinafter referred to as "DBD" usually comprises two zinc finger elements and recognizes a specific Hormone Responsive Element hereinafter referred to as "HRE"

within the promoters of responsive genes. Specific amino acid residues in the "DBD" have been shown to confer DNA sequence binding specificity (Schena, M. & Yamamoto, K.R., Mammalian Glucocorticoid Receptor Derivatives Enhance Transcription in Yeast, Science, 241:965-967, 1988). A Ligand-binding-domain hereinafter referred to as "LBD" is at the carboxy-terminal region of known NRs. In the absence of hormone, the LBD appears to interfere with the interaction of the DBD with its HRE. Hormone binding seems to result in a conformational change in the NR and thus opens this interference (Brzozowski et al., Molecular basis of agonism and antagonism in the oestogen receptor, Nature, 389, 753 – 758, 1997; Wagner et al., A structural role for hormone in the thyroid hormone receptor, Nature, 378, 690 – 697. 1995). A NR without the HBD constitutively activates transcription but at a low level.

Both the amino-terminal domain and the LBD of the NR appear to have transcription activation functions hereinafter referred to as "TAF". Acidic residues in the amino-terminal domains of some nuclear receptors may be important for these transcription factors to interact with RNA polymerase. TAF activity may be dependent on interactions with other protein factors or nuclear components (Diamond et al., Transcription Factor Interactions: Selectors of Positive or Negative Regulation from a Single DNA Element, Science, 249:1266-1272 , 1990). Certain oncoproteins (e.g., c-Jun and c-Fos) can show synergistic or antagonistic activity with glucocorticoid receptors (GR) In transfected cells. Furthermore, the receptors for estrogen and vitamins A and D, and fatty acids have been shown to interact, either physically or functionally, with the Jun and Fos components of AP-1 in the transactivation of steroid- or AP-1 regulated genes.

Coactivators or transcriptional activators are proposed to bridge between sequence specific transcription factors, the basal transcription machinery and in addition to influence the chromatin structure of a target cell. Several proteins like SRC-1, ACTR, and Grip1, which are also cofactors of NRs similar to those disclosed in this invention, interact with NRs in a ligand enhanced manner (Heery et al., A signature motif in transcriptional coactivators mediates binding to nuclear receptors, Nature, 387, 733 – 736; Heinzel et al., A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression, Nature 387, 43 – 47, 1997). Furthermore, the physical interaction with negative receptor-interacting proteins or

corepressors has been demonstrated (Xu et al., Coactivator and Corepressor complexes in nuclear receptor function, Curr Opin Genet Dev, 9 (2), 140 – 147, 1999).

Nuclear receptor modulators like steroid hormones affect the growth and function of specific cells by binding to intracellular receptors and forming nuclear receptor-ligand complexes. Nuclear receptor-hormone complexes then interact with a hormone response element (HRE) in the control region of specific genes and alter specific gene expression.

Over the past decade, new members of the nuclear hormone gene family have been identified that lack known ligands. These orphan receptors can be used to uncover signaling molecules that regulate yet unidentified physiological networks. Some of these orphan receptors are constitutively active and transactivate target genes without the need to interact with a ligand (Mangelsdorf et al., 1995).

The present invention relates to the identification of novel cofactors of the pregnane X receptor (hereinafter referred to as PXR). Note that the human PXR is sometimes also referred to as SXR, for simplicity, we will solely use the name PXR, also for the human protein or gene. PXR is a recently identified orphan nuclear receptor that combines features of nuclear receptors of both the steroid and nonsteroid subfamily: like nonsteroid receptors, PXR binds as a heterodimer with RXR to the HREs of the respective hormone responsive genes. However, interestingly PXR is effectively activated by several steroids, including the naturally occuring pregnanes as well as synthetic glucocorticoids and antiglucocorticoids (Kliever et al., *Cell* 92, 73 (1998)). PXR is abundantly expressed in only a small subset of tissues, predominantly in liver and intestine. Evidence has been provided that PXR acts as a key transcriptional regulator of the genes for cytochrome P450 (CYP) monooxygenases of the 3A subfamily (Moore et al., *PNAS* 97, 7500 (2000); Kliever et al., *Cell* 92, 73 (1998). These hepatic monooxygenases metabolise steroid hormones, including corticosteroids, progestins and androgens as well as a whole range of drugs and xenobiotics. PXR binds to the *CYP3* promoter and the transactivating function of PXR is activated by a range of xenobiotics known to induce CYP3 expression, including rifampicin, RU486, phenoarbital, and pregnenolone. Notably, it has been

shown recently that PXR is activated very efficiently by hyperforin, a constituent of the herbal remedy St. John's wort, which is widely used for the treatment of depression (Moore et al. PNAS 97, 7500 (2000)).

At present it appears that PXR is involved in a novel steroid hormone signaling pathway with implications in the regulation of steroid hormone and sterol homeostasis. Therefore the identification of PXR cofactor proteins that might mediate PXR transactivation activity could provide means for the treatment of numerous diseases or pathophysiological symptoms as a consequence of endocrine malfunction. Furthermore, the examination of PXR variants as well as variants of the PXR interacting proteins could provide valuable clues with respect to the ability of a person to metabolise a certain drug. The kowledge of the genetic background of a person could also help to predict drug-drug interactions.

The present invention provides novel proteins, nucleic acids, and methods useful for developing and identifying compounds for the treatment of such diseases and disorders as metabolic disorders, immunological indications, hormonal dysfunctions and/or neurosystemic diseases and others not specifically mentioned here.

In preferred embodiments of the invention methods are disclosed for testing whether certain compounds promote the interaction of the newly disclosed PXR cofactor proteins with PXR, allowing conclusions on the effects of the compound on PXR activity, and thus on the induction of proteins important for the degradation of xenobiotics, such as the CYP3 proteins.

These novel proteins interact in vivo with the pregnane x receptor and shall hereinafter collectively be referred to as "cofactors".

Identified and disclosed herein are protein sequences for novel cofactors and the nucleic acid sequences encoding these cofactors, which we call: CF1, CF2, CF3, CF4 and CF 44, or simply collectively „CFs".

The importance of this invention is manifested in the effects of the CFs to modulate genes involved in cellular functions like regulation of metabolism and cell

homeostasis, cell proliferation and differentiation, pathological cellular aberrations, or cellular defense mechanisms.

Thus, the CF proteins are useful for screening for pregnane x receptor, thereby providing for agents which may influence the activety of PXR as well as in further preferred embodiments of the invention RXR and thus thereby transcriptionally induced P450 CYP mono oxygenases of the 3A subfamily.

In one aspect of the present invention, we provide isolated nucleic acid sequences for novel CFs. In particular, we provide the cDNA sequences encoding the human CFs.

These nucleic acid sequences have a variety of uses. For example, they are useful for making vectors and for transforming cells, both of which are ultimately useful for production of the CF proteins.

They are also useful as scientific research tools for developing nucleic acid probes for determining expression levels of the cofactor genes, e.g., to identify diseased or otherwise abnormal states. They are useful for developing analytical tools such as anti sense oligonucleotides for selectively inhibiting expression of the cofactor genes to determine physiological responses.

In another aspect of the present invention, we provide a homogenous composition comprising the cofactor proteins. The protein is useful for screening drugs for agonist and antagonist activity, and, therefore, for screening for drugs useful in regulating physiological responses associated with the cofactors according to the Invention. Specifically, antagonists to the CFs could be used to treat metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases. The proteins are also useful for developing antibodies for detection of the proteins.

Flowing from the foregoing are a number of other aspects of the invention, including (a) vectors, such as plasmids, comprising the cofactor nucleic acid sequences that may further comprise additional regulatory elements, e.g., promoters, (b) transformed cells that express the cofactors, (c) nucleic acid probes, (d) antisense

oligonucleotides, (e) agonists, (f) antagonists, and (g) transgenic mammals. Further aspects of the invention comprise methods for making and using the foregoing compounds and compositions.

The foregoing merely summarizes certain aspects of the present invention and is not intended, nor should it be construed, to limit the invention in any manner. All patents and other publications recited herein are hereby incorporated by reference in their entirety.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

THE CF1, CF2, CF3, CF4 AND CF 44 PROTEINS AND NUCLEIC ACIDS:

The present invention comprises, in part, novel CF cofactors of PXR. Particularly preferred embodiments of these cofactors are those having an amino acid sequence substantially the same as SEQ ID NO. 3 for CF1, SEQ ID NO. 6 for CF2, SEQ ID NO. 9 for CF3, SEQ ID NO. 12 for CF 4 and SEQ ID NO. 28 for CF 44 (see also Figures).

As used herein, if reference to the cofactor is made or the cofactor "X",wherein "X" stands for the number designating the cofactor, it is meant as a reference to any protein having an amino acid sequence substantially the same as SEQ ID NO. 3 for CF1, SEQ ID NO. 6 for CF2, SEQ ID NO. 9 for CF3, SEQ ID NO. 12 for CF 4 and SEQ ID NO. 30 for CF 44 (see also Figures).

The present invention also comprises the nucleic acid sequences encoding the cofactors 1 to 4 which nucleic acid sequences are substantially the same as SEQ ID NO. 1 for CF1, SEQ ID NO. 4 for CF2, SEQ ID NO. 7 for CF3, SEQ ID NO. 10 for CF 4 and SEQ ID NO. 28 for CF 44 (see also Figures) all encoding human cofactors as preferred embodiments and/or the complements thereof as shown in SEQ ID NO. 2 for CF1, SEQ ID NO. 5 for CF2, SEQ ID NO. 8 for CF3, SEQ ID NO. 11 for CF 4 and SEQ ID NO. 29 for CF 44 (see also Figures).

Herein the "complement" refers to the complementary strand of the nucleic acid according to the invention, thus the strand that would hybridize to the nucleic acid according to the invention. In accordance with standard biological terminology all DNA sequences herein are however written in 5'-3' orientation, thus the complements depicted (see also figures) are actually "reverse" complements (as also stated in the figures). For simplification purposes they are however some times referred to simply as "complements". One skilled in the art, given the DNA sequence(s) would be able to create the correct reverse complement.

As used herein, a protein "having an amino acid sequence substantially the same as SEQ ID NO x" (where "x" is the number of one of the protein sequences recited in the Sequence Listing) means a protein whose amino acid sequence is the same as SEQ ID NO x or differs only in a way such that at least 50% of the residues compared in a sequence alignment with SEQ ID NO. x are identical, preferably 75% of the residues are identical, even more preferably 95% of the residues are identical and most preferably at least 98% of the residues are identical

Those skilled in the art will appreciate that conservative substitutions of amino acids can be made without significantly diminishing the protein's affinity for interacting proteins, DNA binding sites, cofactor modulators, e.g. small molecular hydrophobic compounds, or RNA.

Other substitutions may be made that increase the proteins' affinity for these compounds. Making and identifying such proteins is a routine matter given the teachings herein, and can be accomplished, for example, by altering the nucleic acid sequence encoding the protein (as disclosed herein), inserting it into a vector, transforming a cell, expressing the nucleic acid sequence, and measuring the binding affinity of the resulting protein, all as taught herein.

As used herein the term "a molecule having a nucleotide sequence substantially the same as SEQ ID NO y" means a nucleic acid encoding a protein "having an amino acid sequence substantially the same as SEQ ID NO y" as defined above. This definition is intended to encompass natural allelic variations in the CF sequences. Cloned nucleic acid provided by the present invention may encode CF proteins of

any species of origin, including (but not limited to), for example, mouse, rat, rabbit, hamster, cat, dog, pig, primate, and human.

Preferably the nucleic acids provided by the invention encode CFs of mammalian, preferably mouse and most preferably human origin.

IDENTIFICATION OF VARIANTS AND HOMOLOGUES AS WELL AS USE OF PROBES:

Nucleic acid hybridization probes provided by the invention are nucleic acids consisting essentially of the nucleotide sequences complementary to any sequence depicted in SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and SEQ ID NO. 28 and/or the complemets thereof as shown in SEQ ID NO. 2, SEQ ID NO. 5, SEQ ID NO. 8, SEQ ID NO. 11 and SEQ ID NO. 29, or parts thereof which are effective in nucleic acid hybridization

Nucleic acid hybridization probes provided by the invention are nucleic acids capable of detecting i.e. hybridizing to the gene encoding the polypeptides according to SEQ ID NO. s: 3, 6, 9, 12 and 30.

Nucleic acid probes are useful for detecting CF gene expression in cells and tissues using techniques well-known in the art, including, but not limited to, Northern blot hybridization, in situ hybridization, and Southern hybridization to reverse transcriptase - polymerase chain reaction product DNAs. The probes provided by the present invention, including oligonucleotide probes derived therefrom, are also useful for Southern hybridization of mammalian, preferably human, genomic DNA for screening for restriction fragment length polymorphism (RFLP) associated with certain genetic disorders. As used herein, the term complementary means a nucleic acid having a sequence that is sufficiently complementary in the Watson-Crick sense to a target nucleic acid to bind to the target under physiological conditions or experimental conditions those skilled in the art routinely use when employing probes.

It is understood in the art that a nucleic acid sequence will hybridize with a complementary nucleic acid sequence under high stringent conditions as defined herein, even though some mismatches may be present. Such closely matched, but not perfectly complementary sequences are also encompassed by the present invention. For example, differences may occur through genetic code degeneracy, or by naturally occurring or man made mutations and such mismatched sequences would still be encompassed by the present claimed invention.

Preferably, the nucleotide sequences of the nuclear cofactors SEQ ID NOs:1, 4, 7, 10 and 28 and/or their complements SEQ ID NO.s 2, 5, 8, 11 and 29 can be used to derive oligonucleotide fragments (probes) of various length. Stretches of 17 to 30 nucleotides are used frequently but depending on the screening parameters longer sequences as 40, 50, 100, 150 up to the full length of the sequence may be used. Those probes can be synthesized chemically and are obtained readily from commercial oligonucleotide providers. Chemical synthesis has improved over the years and chemical synthesis of oligonucleotides as long as 100-200 bases is possible. The field might advance further to allow chemical synthesis of even longer fragments. Alternatively, probes can also be obtained by biochemical *de novo* synthesis of single stranded DNA. In this case the nucleotide sequence of the nuclear receptors or their complements serve as a template and the corresponding complementary strand is synthesized. A variety of standard techniques such as nick translation or primer extension from specific primers or short random oligonucleotides can be used to synthesize the probe (Sambrook, J., Fritsch, E.F. & Maniatis , T. Molecular cloning: a laboratory manual. Cold Spring Harbor Press, Cold Spring Harbor, 1989)). Nucleic acid reproduction technologies exemplified by the polymerase chain reaction (Saiki, R.K. et *al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491 (1988)) are commonly applied to synthesize probes. In the case of techniques using specific primers the nucleic acid sequences of the nuclear receptors or their complements are not only used as a template in the biochemical reaction but also to derive the specific primers which are needed to prime the reaction.

In some cases one might also consider to use the nucleic acid sequence of the cofactors or their complements as a template to synthesize an RNA probe. A promoter sequence for a DNA-dependent RNA polymerase has to be introduced at the 5'-end of sequence. As an example this can be done by cloning the sequence in a vector which carries the respective promoter sequence. It is also possible to introduce the needed sequence by synthesizing a primer with the needed promoter in the form of a 5' "tail". The chemical synthesis of a RNA probe is another option.

Appropriate means are available to detect the event of a hybridization. There is a wide variety of labels and detection systems, e.g. radioactive isotopes, fluorescent, or chemiluminescent molecules which can be linked to the probe. Furthermore, there are methods of introducing haptens which can be detected by antibodies or other ligands such as the avidin/biotin high affinity binding system.

Hybridization can take place in solution or on solid phase or in combinations of the two, e.g. hybridization in solution and subsequent capture of the hybridization product onto a solid phase by immobilized antibodies or by ligand coated magnetic beads.

Hybridization probes act by forming selectively duplex molecules with complementary stretches of a sequence of a gene or a cDNA. The selectivity of the process can be controlled by varying the conditions of hybridization. To select sequences which are identical highly homologous to the sequence of interest stringent conditions for the hybridization will be used, e.g. low salt in the range of 0.02 M to 0.15 M salt and/or high temperatures in the range from 50°C degrees centigrade to 70°C degrees centigrade. Stringency can be further improved by the addition of formamide to the hybridisation solution. The use of stringent conditions which means that only little mismatch or a complete match will lead to a hybridization product would be used to isolate closely related members of the same gene family. Thus, as used herein stringent hybridization conditions are those where between 0.02 M to 0.15 M salt and/or high temperatures in the range from 50°C degrees centigrade to 70°C degrees centigrade are applied.

The use of highly stringent conditions or conditions of "high stringency" means that only very little mismatch or a complete match which lead to a hybridization product

would be used to isolate very closely related members of the same gene family. Thus, as used herein highly stringent hybridization conditions are those where between 0.02 – 0.3 M salt and 65°C degrees centigrade are applied for about 5 to 18 hours of hybridization time and additionally, the sample filters are washed twice for about 15 minutes each at between 60°C – 65°C degrees centigrade, wherein the first washing fluid contains about 0.1 M salt (NaCl and/or Sodium Citrate) and the second contains only about 0.02 M salt (NaCl and/or Sodium Citrate). In a preferred embodiment the following conditions are considered to be highly stringent:

Hybridisation in a buffer containing 2 x SSC (0.03 M Sodium Citrate, 0.3 M NaCl) at 65°C – 68°C degrees centigrade for 12 hours, followed by a washing step for 15 minutes in 0.5 x SSC, 0.1% SDS, and a washing step for 15 minutes at 65°C degrees centigrade in 0.1 x SSC, 0.1% SDS.

Less stringent hybridization conditions, e.g. 0.15 M salt - 1 M salt and/or temperatures from 22°C degrees centigrade to 56°C degrees centigrade are applied in order to detect functionally equivalent genes in the same species or for orthologous sequences from other species.

Unspecific hybridization products are removed by washing the reaction products repeatedly in 2 x SSC solution and increasing the temperature.

DEGENERATE PCR AND CLONING OF HOMOLOGUES

The nucleotide sequences of the cofactors CF1 to CF4 or their complements can be used to design primers for a polymerase chain reaction. Due to the degeneracy of the genetic code the respective amino acid sequence is used to design oligonucleotides in which varying bases coding for the same amino acid are included. Numerous design rules for degenerate primers have been published (Compton et al, 1990). As in hybridization there are a number of factors known to vary the stringency of the PCR. The most important parameter is the annealing temperature. To allow annealing of primers with imperfect matches annealing temperatures are often much lower than the standard annealing temperature of 55°C, e.g. 35°C to 52°C degrees

can be chosen. PCR reaction products can be cloned. Either the PCR product is cloned directly, with reagents and protocols from commercial manufacturers (*e.g.* from Invitrogen, San Diego, USA). Alternatively, restriction sites can be introduced intro the PCR product via a 5'-tail of the PCR primers and used for cloning.

GENETIC VARIANTS

Fragments from the nucleotide sequence of the cofactors or their complements can be used to cover the whole sequence with overlapping sets of PCR primers. These primers are used to produce PCR products using genomic DNA from a human diversity panel of healthy individuals or genomic DNA from individuals which are phenotypically conspicuous. The PCR products can be screened for polymorphisms, for example by denaturing gradient gel electrophoresis, binding to proteins detecting mismatches or cleaving heteroduplices or by denaturing high-performance liquid chromatography. Products which display mutations need to be sequenced to identify the nature of the mutation. Alternatively, PCR products can be sequenced directly omitting the mutation screening step to identify genetic polymorphisms. If genetic variants are identified and are associated with a discrete phenotype, these genetic variations can be included in diagnostic assays. The normal variation of the human population is of interest in designing screening assays as some variants might interact better or worse with a respective lead, i.e. therapeutic or potentially therapeutic substance (a pharmacodynamic application). Polymorphisms or mutations which can be correlated to phenotypic outcome are a tool to extend the knowledge and the commercial applicability of the nucleotide sequences of the claimed cofactors or their complements or their gene products, as variants might have a slightly different molecular behavior or desired properties. Disease-causing mutations or polymorphisms allow the replacement of this disease inducing gene copy with a wild-type copy by means of gene therapy approaches and/or the modulation of the activity of the gene product by drugs. Disease-causing mutations or polymorphisms in the CF cofactors allow predictions on the induction of the CYP3 genes by certain substances, and thus on the degradation of drugs in the body.

PREPARATION OF POLYNUCLEOTIDES:

DNA which encodes the cofactor may be obtained, in view of the instant disclosure, by chemical synthesis, by screening reverse transcripts of mRNA from appropriate cells or cell line cultures, by screening genomic libraries from appropriate cells, or by combinations of these procedures, as illustrated below.

Screening of mRNA or genomic DNA may be carried out with oligonucleotide probes generated from the CF nucleotide sequences information provided herein.

Probes may be labeled with a detectable group such as a fluorescent group, a radioactive atom or a chemiluminescent group in accordance with known procedures and used in conventional hybridization assays, as described in greater detail in the Examples below. Alternatively, the CF nucleotide sequences may be obtained by use of the polymerase chain reaction (PCR) procedure, with the PCR oligonucleotide primers being produced from the CF nucleotide sequences provided herein, according to SEQ ID NO 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and SEQ ID NO. 28 and/or the complements thereof as shown in SEQ ID NO. 2, SEQ ID NO. 5, SEQ ID NO. 8, SEQ ID NO. 11 and SEQ ID NO. 29, or parts thereof.

Upon purification or synthesis, the nucleic acid according to the invention may be labeled, e.g. for use as a probe.

As single and differential labeling agents and methods, any agents and methods which are known in the art can be used. For example, single and differential labels may consist of the group comprising enzymes such as β-galactosidase, alkaline phosphatase and peroxidase, enzyme substrates, coenzymes, dyes, chromophores, fluorescent, chemiluminescent and bioluminescent labels such as FITC, Cy5, Cy5.5, Cy7, Texas-Red and IRD40(Chen et al. (1993), J. Chromatog. A 652: 355-360 and Kambara et al. (1992), Electrophoresis 13: 542-546), ligands or haptens such as biotin, and radioactive isotopes such as ^{3}H, ^{35}S, ^{32}P ^{125}I and ^{14}C.

EXPRESSION OF THE CF PROTEIN/POLYPETIDES:

The CF nucleic acids or polypeptides may be synthesized in host cells transformed with a recombinant expression construct comprising a nucleic acid encoding any of

the cofactors according to the invention. Such a recombinant expression construct can also be comprised of a vector that is a replicable DNA construct.

Amplification vectors do not require expression control domains. All that is needed is the ability to replicate in a host, usually conferred by an origin of replication, and a selection gene to facilitate recognition of transformants. See, Sambrook et al., Molecular Cloning: A Laboratory Manual (2nd Edition, Cold Spring Harbor Press, New York, 1989).

An expression vector comprises a polynucleotide operatively linked to a prokaryotic promoter. Alternatively, an expression vector is a polynucleotide operatively linked to an enhancer-promoter that is a eukaryotic promoter, and the expression vector further has a polyadenylation signal that is positioned 3' of the carboxy-terminal amino acid and within a transcriptional unit of the encoded polypeptide. A promoter is a region of a DNA molecule typically within about 500 nucleotide pairs in front of (upstream of) the point at which transcription begins (*i.e.*, a transcription start site). In general, a vector contains a replicon and control sequences which are derived from species compatible with the host cell. The vector ordinarily carries a replication site, as well as marking sequences which are capable of providing phenotypic selection in transformed cells.

Another type of discrete transcription regulatory sequence element is an enhancer. An enhancer provides specificity of time, location and expression level for a particular encoding region (e.g., gene). A major function of an enhancer is to increase the level of transcription of a coding sequence in a cell.

As used herein, the phrase "enhancer-promoter" means a composite unit that contains both enhancer and promoter elements. An enhancer-promoter is operatively linked to a coding sequence that encodes at least one gene product.

An enhancer-promoter used in a vector construct of the present invention may be any enhancer-promoter that drives expression in a prokaryotic or eukaryotic cell to be transformed/transfected.

A coding sequence of an expression vector is operatively linked to a transcription terminating region. RNA polymerase transcribes an encoding DNA sequence through a site where polyadenylation occurs.

An expression vector that comprises a polynucleotide that encodes polypeptides of the cofactors. Such a polynucleotide is meant to include a sequence of nucleotide bases encoding a CF polypeptide sufficient in length to distinguish said segment from a polynucleotide segment encoding a non- cofactor polypeptide.

A polypeptide of the invention may also encode biologically functional polypeptides or peptides which have variant amino acid sequences, such as with changes selected based on considerations such as the relative hydropathic score of the amino acids being exchanged.

These variant sequences are those isolated from natural sources or induced in the sequences disclosed herein using a mutagenic procedure such as site-directed mutagenesis.

Furthermore, an expression vector of the present invention may contain regulatory elements for optimized translation of the polypeptide in prokaryotic or eukaryotic systems. These sequences are operatively located around the transcription start site and are most likely similar to ribosome recognition sites like prokaryotic ribosome binding sites (RBS) or eukaryotic Kozak sequences as known in the art (Kozak M., Initiation of translation in prokaryotes and eukaryotes. *Gene* 234, 187-208 (1999)).

An expression vector of the present invention is useful both as a means for preparing quantities of the CFs' polypeptide-encoding DNA itself, and as a means for preparing the encoded CFs' polypeptide and peptides. It is contemplated that where cofactor polypeptides of the invention are made by recombinant means, one may employ either prokaryotic or eukaryotic expression vectors as shuttle systems.

Where expression of recombinant CF1, CF2, CF3, CF4 or CF30 polypeptide is desired and a eukaryotic host is contemplated, it is most desirable to employ a vector such as a plasmid, that incorporates a eukaryotic origin of replication. Additionally, for

the purposes of expression in eukaryotic systems, one desires to position the cofactor encoding sequence or if desired parts thereof adjacent to and under the control of an effective eukaryotic promoter. To bring a coding sequence under control of a promoter, whether it is eukaryotic or prokaryotic, what is generally needed is to position the 5' end of the translation initiation side of the proper translational reading frame of the polypeptide between about 1 and about 2000 nucleotides 3' of or downstream with respect to the promoter chosen.

Furthermore, where eukaryotic expression is anticipated, one would typically desire to incorporate into the transcriptional unit which includes the CF polypeptide, an appropriate polyadenylation side.

The invention provides homogeneous compositions of mammalian cofactor polypeptides produced by transformed prokaryotic or eukaryotic cells as provided herein. Such homogeneous compositions are intended to be comprised of mammalian cofactor protein that comprises at least 90% of the protein in such homogenous composition. The invention also provides membrane preparation from cells expressing the mammalian cofactors polypeptides as the result of transformation with a recombinant expression construct, as described here.

Within the scope of the present invention the terms recombinant protein or coding sequence both also include tagged versions of the proteins depicted in SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12 and SEQ ID NO. 30 and/or and fusion proteins of said proteins with any other recombinant protein. Tagged versions here means that small epitopes of 3-20 amino acids are added to the original protein by extending the coding sequence either at the 5´or the 3´terminus leading to N-terminal or C-terminal extended proteins respectively, or that such small epitopes are included elsewhere in the protein. The same applies for fusion proteins where the added sequences are coding for longer proteins, varying between 2 and 100 kDa. Tags and fusion proteins are usually used to facilitate purification of recombinant proteins by specific antibodies or affinity matrices or to increase solubility of recombinant proteins within the expression host. Fusion proteins are also of major use as essential parts of yeast two hybrid screens for interaction partners of recombinant proteins.

Tags used in the scope of the present invention may include but are not limited to the following: EEF (alpha Tubulin), B-tag (QYPALT), E tag (GAPVPYPDPLEPR) c-myc Tag (EQKLISEEDL), Flag epitope (DYKDDDDK, HA tag (YPYDVPDYA), 6 or 10 x His Tag, HSV (QPELAPEDPED), Pk-Tag (GKPIPNPLLGLDST), protein C (EDQVDPRLIDGK), T7 (MASMTGGQQMG), VSV-G (YTDIEMNRLGK), Fusion proteines may include Thioredoxin, Glutathiontransferase (GST), Maltose binding Protein (MBP), Cellulose Binding protein, calmodulin binding protein, chitin binding protein, ubiquitin, the Fc part of Immunoglobulins, and the IgG binding domain of Staphylococcus aureus protein A. These examples of course are illustrative and not limiting.

For expression of recombinant proteins in living cells or organisms, vector constructs harboring recombinant cofactors as set forth in SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and SEQ ID NO. 28 are transformed or transfected into appropriate host cells. Preferably, a recombinant host cell of the present invention is transfected with a polynucleotide of SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and SEQ ID NO. 28.

Means of transforming or transfecting cells with exogenous polynucleotide such as DNA molecules are well known in the art and include techniques such as calcium-phosphate- or DEAE-dextran-mediated transfection, protoplast fusion, electroporation, liposome mediated transfection, direct microinjection and virus infection (Sambrook et al., 1989).

The most frequently applied technique for transformation of prokaryotic cells is transformation of bacterial cells after treatment with Calciumchloride to increase permeability (Dagert & Ehrlich, 1979), but a variety of other methods is also available for one skilled in the art.

The most widely used method for transfection of eukaryotic cells is transfection mediated by either calcium phosphate or DEAE-dextran. Although the mechanism remains obscure, it is believed that the transfected DNA enters the cytoplasm of the cell by endocytosis and is transported to the nucleus. Depending on the cell type, up

to 90% of a population of cultured cells may be transfected at any one time. Because of its high efficiency, transfection mediated by calcium phosphate or DEAE-dextran is the method of choice for studies requiring transient expression of the foreign nucleic acid in large numbers of cells. Calcium phosphate-mediated transfection is also used to establish cell lines that integrate copies of the foreign DNA, which are usually arranged in head-to-tail tandem arrays into the host cell genome.

In the protoplast fusion method, protoplasts derived from bacteria carrying high numbers of copies of a plasmid of interest are mixed directly with cultured mammalian cells. After fusion of the cell membranes (usually with polyethylene glycol), the contents of the bacterium are delivered into the cytoplasm of the mammalian cells and the plasmid DNA is transported to the nucleus. Protoplast fusion is not as efficient as transfection for many of the cell lines that are commonly used for transient expression assays, but it is useful for cell lines in which endocytosis of DNA occurs inefficiently. Protoplast fusion frequently yields multiple copies of the plasmid DNA tandemly integrated into the host chromosome.

The application of brief, high-voltage electric pulses to a variety of mammalian and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores. Electroporation may be extremely efficient and may be used both for transient expression of cloned genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

Liposome transfection involves encapsulation of DNA and RNA within liposomes, followed by fusion of the liposomes with the cell membrane. The mechanism of how DNA is delivered into the cell is unclear but transfection efficiencies may be as high as 90%.

Direct microinjection of a DNA molecule into nuclei has the advantage of not exposing DNA to cellular compartments such as low-pH endosomes. Microinjection

is therefore used primarily as a method to establish lines of cells that carry integrated copies of the DNA of interest.

The use of adenovirus as a vector for cell transfection is well known in the art. Adenovirus vector-mediated cell transfection has been reported for various cells (Stratford-Perricaudet et al., 1992).

A transfected cell may be prokaryotic or eukaryotic, transfection may be transient or stable. Where it is of interest to produce a full length human CF1, CF2, CF3, CF4 or CF 44 protein, cultured mammalian or human cells are of particular interest.

In another aspect, the recombinant host cells of the present invention are prokaryotic host cells. In addition to prokaryotes, eukaryotic microbes, such as yeast may also be used illustrative examples for suitable cells and organisms for expression of recombinant proteins are belonging to but not limited to the following examples: Insect cells, such as Drosophila Sf21, SF9 cells or others, Expression strains of Escherichia coli, such as XL1 blue, BRL21, M15, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Hansenlua polymorpha and Pichia pastoris strains, immortalized mammalian cell lines such as AtT-20, VERO and HeLa cells, Chinese hamster ovary (CHO) cell lines, and W138, BHK, COSM6, COS-7, 293 and MDCK cells, BHK-21 cells, Att 20HeLa cells, HeK 294, T47 D cells and others.

Expression of recombinant proteins within the scope of this invention can also be performed in vitro. This may occur by a two step procedure, thereby producing first mRNA by in vitro transcription of an apt polynucleotide construct followed by in vitro translation with convenient cellular extracts. These cellular extracts may be reticulocyte lysates but are not limited to this type. In vitro transcription may be performed by T7 or SP6 DNA polymerase or any other RNA polymerase which can recognize per se or with the help of accessory factors the promoter sequence contained in the recombinant DNA construct of choice. Alternatively one of the recently made available one step coupled transkription/translation systems may be used for in vitro translation of DNA coding for the proteins of this invention. One illustrative but not limiting example for such a system is the TNT® T7 Quick System by Promega.

Expression of recombinant proteins in transfected cells may occur constitutively or upon induction. Procedures depend on the Cell/vector combination used and are well known in the art.

In all cases, transfected cells are maintained for a period of time sufficient for expression of the recombinant cofactor proteins according to the invention. A suitable maintenance time depends strongly on the cell type and organism used and is easily ascertainable by one skilled in the art. Typically, maintenance time is from about 2 hours to about 14 days. For the same reasons and for sake of protein stability and solubility incubation temperatures during maintenance time may vary from 20°C to 42 °C.

Recombinant proteins are recovered or collected either from the transfected cells or the medium in which those cells are cultured. Recovery comprises cell disruption, isolation and purification of the recombinant protein. Isolation and purification techniques for polypeptides are well-known in the art and include such procedures as precipitation, filtration, chromatography, electrophoresis and the like.

In a preferred embodiment, purification includes but is not limited to affinity purification of tagged or nontagged recombinant proteins. This is a well established robust technique easily adapted to any tagged protein by one skilled in the art. For affinity purification of tagged proteines, small molecules such as gluthathione, maltose or chitin, specific proteins such as the IgG binding domain of Staphylococcus aureus protein A, antibodies or specific chelates which bind with high affinity to the tag of the recombinant protein are employed. For affinity purification of non-tagged proteins specific monoclonal or polyclonal antibodies, which were raised against said protein, can be used. Alternatively immobilized specific interactors of said protein may be employed for affinity purification. Interactors include native or recombinant proteins as well as native or artificial specific low molecular weight ligands.

CHEMICAL SYNTHESIS OF THE POLYPEPTIDE ACCORDING TO THE INVENTION:

Alternatively, the protein itself may be produced using chemical methods to synthesize any of the amino acid sequences according to the invention (SEQ ID No: SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12 and/or SEQ ID NO. 30) or that is encoded by the nucleotide sequences according to the invention (SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and/or SEQ ID NO. 28) or a portion thereof. For example, peptide synthesis can be performed using conventional Merrifield solid phase f-Moc or t-Boc chemistry or various solid-phase techniques (Roberge, J. Y. et al. (1995) Science 269: 202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer). The newly synthesized peptide(s) may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, N.Y.). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.,* the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequences according to the invention, *i.e.* SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12 and/or SEQ ID NO. 30 or the sequence that is encoded by SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and/or SEQ ID NO. 28 or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

COMPLEXES OF THE COFACTORS ACCORDING TO THE INVENTION WITH OTHER POLYPETIDES

As outlined above CFs all bind the pregnane X receptor in vivo. In a preferred embodiment of the invention the CFs are complexed with this polypeptide or a portion thereof as disclosed in SEQ ID NO. 15 and/or 18. Such complexes are particular suited for all forms of binding or screening assays (see also below). Thus, in a preferred embodiment of the invention such assays are performed with complexes of the pregnane X receptor associated with one or more of the CF cofactors.

Such a complex may additionally comprise other cofactors such as RXR. In one embodiment of the invention a heterotrimeric complex of PXR, RXR and CF1, CF2,

CF3 or CF4 is claimed. RXR herein refers equally to the alpha, beta or gamma form as encoded by SEQ ID NOs. 19, 22 and 25 or depicted in SEQ ID NOs. 21, 24 and 27 or any portion thereof.

Such heterotrimers and multimers may be used in binding and screening assays as outlined below.

In one embodiment of the invention the entire CF polypeptide is part of the complex but only a portion, e.g. a truncated fragment of the other polypeptide (PXR or RXR) is part of the complex.

SCREENING ASSAYS

In still a further embodiment, the present invention concerns a method for identifying new inhibitory or stimulatory substances of the cofactors according to the invention, these substances may be termed as "candidate substances". It is contemplated that this screening technique proves useful in the general identification of compounds that serve the purpose of inhibiting or stimulating cofactor activity.

In one embodiment of the invention the following substances are disclosed as interactors of the PXR cofactor complexes:

Steroids: dexamethasone-t-butylacetate, RU486, progesterone, 17-alpha-hydroxyprogesterone, 1,16-alpha dimethylpregnenolone, 17-alpha-hydroxypregnenonlone, pregnenonlone, 5beta-pregnane-3,20-dione, pregnenonlone-16-carbonitrile, 5beta-pregnane-3,20-dione, androstanol, corticosterone, dehydroeplandrosterone, dihydroxytestosterone, estradiol, cortisol, cortisone, dihydroxytestosterone.

Other substances: transnonachlor, chlordane, spironolactone, cyproterone acetate, rifampicin, nefipine, diethylstilbestrol, coumesterol, clotrimazole, lovastatin, phenoarbital, pthalic acid, nonylphenol, 1,4-bis(2-(3,5-dichloropyridyloxy1))benzene,

One may use the screening method to identify such substances which activate the transactivation function of PXR, and thus lead to elevated expression of CYP3A genes. In turn, this will alter the ability of an individual treated with the substance to degrade xenobiotics, including drugs.

This also includes the use of heteromultimeric complexes of the cofactor protein with other proteins, such as the pregnane x receptor protein, or any other binding partner.

Accordingly, in screening assays to identify pharmaceuticals agents which affect cofactor acitivity, it is proposed that compounds isolated from natural sources, such as fungal extracts, plant extracts, bacterial extracts, higher eukaryotic cell extracts, or even extracts from animal sources, or marine, forest or soil samples, may be assayed for the presence of potentially useful pharmaceutical agents.

It will be understood that that the pharmaceutical agents to be screened can also be derived from chemical compositions or man-made compounds. The candidate substances can could also include monoclonal or polyclonal antibodies, peptides or proteins, such as those derived from recombinant DNA technology or by other means, including chemical peptide synthesis. The active compounds may include fragments or parts or derivatives of naturally-occurring compounds or may be only found as active combinations of known compounds which are otherwise inactive. We anticipate that such screens will in some cases lead to the isolation of agonists of nuclear receptors or cofactors, in other cases to the isolation of antagonists. In other instances, substances will be identified that have mixed agonistic and antagonistic effects, or affect nuclear receptors or cofactors in any other way.

In another embodiment, the invention concerns the isolation of substance inhibiting the interaction of the cofactor protein and the pregnane x receptor. Such substances are useful for the development of drugs against diseases such as metabolic disorders, immunological indications, hormonal dysfunctions and/or neurosystemic diseases ad diseases related to a different ability to degrade xenobiotic substances or related to defects in steroid homeostasis. Substances disrupting the interactions may be isolated by a variety of screening methods including the two hybrid system or the reverse two hybrid system (Lenna C.A. and Hannink, M. 1996, Nucl. Acids Res.

24: 3341-3347), or any variation of cellular or cell free assays as described in this invention, as is obvious to anyone skilled in the art.

In an important embodiment of the invention, the binding of the cofactor protein and the pregnane x receptor can be used to monitor the binding of a substance to one of the binding partners. The substance, which can be a small molecule such as a ligand to a nuclear receptor, will lead to a change in the allosteric conformation of the binding protein which in consequence leads to a loss of the interaction of the two proteins. Using this effect of ligand-dependent protein-protein interactions one can design assays where the protein-protein interaction serves as a surrogate read-out for the binding of one of the proteins to small molecule ligand. Any assay method which is useful for the measurement of protein-protein interactions can be used for such an indirect assay. Such assay methods are well known in the art and include the methods described in this patent under "Cell free assays" and "Cell based assays". In a preferred embodiment, this assay will measure the binding of substances to PXR, resulting in an effect on the interaction of PXR with the cofactor.

CELL BASED ASSAYS

To identify a candidate substance capable of influencing the cofactor protein activity, one first obtains a recombinant cell line. One designs the cell line in such a way that the activity of the cofactor leads to the expression of a protein which has an easily detectable phenotype (a reporter), such as luciferase, fluorescent proteins such as green or red fluorescent protein, beta-galactosidase, alpha-galactosidase, beta-lactamase, chloramphenicol-acetyl-transferase, beta-glucuronidase, or any protein which can be detected by a secondary reagent such as an antibody.

Methods for detecting proteins using antibodies, such as ELISA assays, are well known to those skilled in the art.

Here, the amount of reporter protein present reflects the activity of the cofactor. This recombinant cell line is then screened for the effect of substances on the expression of the reporters, thus measuring the effect of these substances on the activity of the cofactor. These substances can be derived from natural sources, such as fungal

extracts, plant extracts, bacterial extracts, higher eukaryotic cell extracts, or even extracts from animal sources, or marine, forest or soil samples, may be assayed for the presence of potentially useful pharmaceutical agents. It will be understood that that the pharmaceutical agents to be screened may be derived from chemical compositions or man-made compounds.

The candidate substances can also include monoclonal or polyclonal antibodies, peptides or proteins, such as those derived from recombinant DNA technology or by other means, including chemical peptide synthesis. The active compounds may include fragments or parts or derivatives of naturally-occurring compounds or may be only found as active combinations of known compounds which are otherwise inactive.

In general the assay can be performed by firstly bringing a suitable cell containing a reporter gene which transcription is influenced by the cofactors activity in contact with a compound and secondly monitoring the expression of the reporter gene to evaluate the effect of the compound on the activity of the cofactor.

In other embodiments of the invention assays are included where measuring the activity of di- or multimeric complexes of the cofactor and other proteins such as PXR or RXR. Further included are assays aiming at the identification of compounds which specifically influence only the monomeric, homodimeric or homomultimeric form of the cofactor, or influencing only multimeric forms of the cofactor. Such assays include measuring the effect of a compound on the cofactor in the absence of a binding partner, and measuring the effect of a compound on the cofactor in the presence of a binding partner, such as PXR. One skilled in the art will find numerous more assays which are equally covered by the invention.

A cell line where the activity of PXR or any other nuclear receptor determines the expression of a reporter can be obtained by generating an artificial promoter upstream of the reporter gene, which contains preferably multiple copies of HREs to which PXR or any other nuclear receptor binds.

Furthermore, transgenic animals described in the invention can be used to derive cell lines useful for cellular screening assays.

Cell lines useful for such an assay include many different kinds of cells, including prokaryotic, animal, fungal, plant and human cells. Yeast cells can be used in this assay, including *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* cells.

One way of building cellular assays is by measuring the effect of compounds is the use of the two hybrid system (see for example see, for example, U.S. Pat. No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Biotechniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; PCT Publication No. WO 94/10300, and U.S. Pat. No. 5,667,973), or possible variants of the basic two hybrid system as discussed e.g in Vidal M, Legrain P, Nucleic Acids Res 1999 Feb 15;27(4):919-29. Briefly, the two hybrid assay relies on reconstituting In vivo a functional transcriptional activator protein from two separate fusion proteins. In particular, the method makes use of chimeric genes which express hybrid proteins. To illustrate, a first hybrid gene comprises the coding sequence for a DNA-binding domain of a transcriptional activator fused in frame to the coding sequence for a cofactor. The second hybrid protein encodes a transcriptional activation domain fused in frame to another gene, for example PXR. If the cofactor and PXR proteins are able to interact, they bring into close proximity the two domains of the transcriptional activator. This proximity is sufficient to cause transcription of a reporter gene which is operably linked to a transcriptional regulatory site responsive to the transcriptional activator, and expression of the reporter gene can be detected and used to score for the interaction of the cofactor and PXR proteins. Suitable host cells for such assays include yeast cells, but also mammalian cells or bacterial cells.

In such assays, one primarily measures the effect of a compound on a given interaction involving the CF cofactors and a binding protein. In a preferred embodiment of the invention systems using other hosts such as prokaryotes as *E. coli*, or eukaryotic mammalian cells are described.

Two hybrid systems using hybrid protein fusions with other proteins than transcription factors, including enzymes such as beta-galactosidase or *dihydrofolate reductase* may also be applied. These assays are useful both to monitor the effect of a compound, including peptides, proteins or nucleic acids on an interaction of a cofactor with a given binding partner, as well as to identify novel proteins or nucleic acids interacting with the cofactor.

CELL-FREE ASSAYS

Recombinant forms of the polypeptides according to SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12 and/or SEQ ID NO. 30 can be used in cell-free screening assays aiming at the isolation of compounds affecting the activity of cofactors. In such an assay, the cofactor polypeptides are brought into contact with a substance to test if the substance has an effect on the activity of the cofactors.

The detection of an interaction between an agent and a cofactor may be accomplished through techniques well-known in the art. These techniques include but are not limited to centrifugation, chromatography, electrophoresis and spectroscopy. The use of isotopically labeled reagents in conjunction with these techniques or alone is also contemplated. Commonly used radioactive isotopes include ^{3}H, ^{14}C, . ^{22}Na, ^{32}P, ^{33}P, ^{35}S, ^{45}Ca, ^{60}Co, ^{125}I, and ^{131}I. Commonly used stable isotopes include ^{2}H, . ^{13}C, ^{15}N, ^{18}O.

For example, if an agent binds to any of the cofactors of the present invention, the binding may be detected by using radiolabeled agent or radiolabeled cofactor. Briefly, if radiolabeled agent or radiolabeled cofactor is utilized, the agent-cofactor complex may be detected by liquid scintillation or by exposure to x-ray film or phosho-imaging devices.

One way to screen for substances affecting cofactor activity is to measure the effect of the substance on the binding affinity of the cofactor to other proteins or molecules, such as activators or repressors, DNA, RNA, other proteins, antibodies peptides or other substances, including chemical compounds known to affect receptor activity or to a nuclear receptor itself. Assays measuring the binding of a protein to a ligand are

well known in the art, such as ELISA assays, FRET assays, bandshift assays, plasmon-resonance based assays, scintillllation proximity assays, fluorescence polarization assays, alpha screen assays.

In one example, a mixture containing a cofactor polypeptide, effector and candidate substance is allowed to incubate. The unbound effector is separable from any effector/cofactor complex so formed. One then simply measures the amount of each (*e.g.*, versus a control to which no candidate substance has been added). This measurement may be made at various time points where velocity data is desired. From this, one determines the ability of the candidate substance to alter or modify the function of the cofactor.

Numerous techniques are known for separating the effector from effector/cofactor complex, and all such methods are intended to fall within the scope of the invention. This includes the use of thin layer chromatographic methods (TLC), HPLC, spectrophotometric, gas chromatographic/mass spectrophotometric or NMR analyses. Another method of separation is to immobilize one of the binding partners on a solid support, and to wash away any unbound material. It is contemplated that any such technique may be employed so long as it is capable of differentiating between the effector and complex, and may be used to determine enzymatic function such as by identifying or quantifying the substrate and product.

A screening assay in which candidate agent binding of cofactors is analysed can include a number of conditions. These conditions include but are not limited to pH, temperature, tonicity, the presence of relevant other proteins, and relevant modifications to the polypeptide such as glycosylation or lipidation. It is contemplated that the cofactors can be expressed and utilized in a prokaryotic or eukaryotic cell. The host cell expressing the cofactors can be used whole or the cofactor can be isolated from the host cell. The cofactor can be membrane bound in the membrane of the host cell or it can be free in the cytosol of the host cell. The host cell can also be fractionated into sub-cellular fractions where the cofactor can be found. For example, cells expressing the cofactor can be fractionated into the nuclei, the *endoplasmic reticulum*, vesicles, or the membrane surfaces of the cell.

pH is preferably from about a value of 6.0 to a value of about 8.0, more preferably from about a value of about 6.8 to a value of about 7.8, and most preferably, about 7.4. In a preferred embodiment, temperature is from about 20°C degrees to about 50°C degrees more preferably, from about 30°C degrees to about 40°C degrees and even more preferably about 37°C degrees. Osmolality is preferably from about 5 milliosmols per liter (mosm/L) to about 400 mosm/l, and more preferably, from about 200 milliosmols per liter to about 400 mosm/l and, even more preferably from about 290 mosm/L to about 310 mosm/L. The presence of further cofactors or other proteins can be required for the proper functioning of the cofactors according to the invention. Typical chemical cofactors include sodium, potassium, calcium, magnesium, and chloride. In addition, small, non-peptide molecules, known as prosthetic groups may also be required. Other biological conditions needed for cofactor function are well-known in the art.

It is well-known in the art that proteins can be reconstituted in artificial membranes, vesicles or liposomes. (Danboldt et al.,1990). The present invention contemplates that the cofactor can be incorporated into artificial membranes, vesicles or liposomes. The reconstituted cofactor can be utilized in screening assays.

It is further contemplated that a cofactor of the present invention can be coupled to a solid support, e.g., to agarose beads, polyacrylamide beads, polyacrylic, sepharose beads or other solid matrices capable of being coupled to polypeptides. Well-known coupling agents include cyanogen bromide (CNBr), carbonyldiimidazole, tosyl chloride, diaminopimelimidate, and glutaraldehyde.

In a typical screening assay for identifying candidate substances, one employs the same recombinant expression host as the starting source for obtaining the cofactor polypeptide, generally prepared in the form of a crude homogenate. Recombinant cells expressing the cofactor are washed and homogenized to prepare a crude polypeptide homogenate in a desirable buffer such as disclosed herein. In a typical assay, an amount of polypeptide from the cell homogenate, is placed into a small volume of an appropriate assay buffer at an appropriate pH. Candidate substances, such as agonists and antagonists, are added to the admixture in convenient

concentrations and the interaction between the candidate substance and the cofactor polypeptide is monitored (see also Fig. 1).

Where one uses an appropriate known substrate for the cofactors, one can, in the foregoing manner, obtain a baseline activity for the recombinantly produced cofactors. Then, to test for inhibitors or modifiers of the cofactor function, one can incorporate into the admixture a candidate substance whose effect on the cofactor is unknown. By comparing reactions which are carried out in the presence or absence of the candidate substance, one can then obtain information regarding the effect of the candidate substance on the normal function of the cofactor.

Accordingly, this aspect of the present invention will provide those of skill in the art with methodology that allows for the identification of candidate substances having the ability to modify the action of cofactor polypeptides in one or more manners.

Additionally, screening assays for the testing of candidate substances are designed to allow the determination of structure-activity relationships of agonists or antagonists with the cofactors, e.g., comparisons of binding between naturally-occurring hormones or other substances capable of interacting with or otherwise modulating the cofactor; or comparison of the activity caused by the binding of such molecules to the cofactor.

In certain aspects, the polypeptides of the invention are crystallized in order to carry out x-ray crystallographic studies as a means of evaluating interactions with candidate substances or other molecules with the cofactor polypeptide. For instance, the purified recombinant polypeptides of the invention, i.e. of the cofactors according to the invention, when crystallized in a suitable form, are amenable to detection of intra-molecular interactions by x-ray crystallography. In another aspect, the structure of the polypeptides can be determined using nuclear magnetic resonance.

PHARMACEUTICAL COMPOSITION:

This invention provides a pharmaceutical composition comprising an effective amount of an agonist or antagonist drug identified by the method described herein

and a pharmaceutically acceptable carrier. Such drugs and carrier can be administered by various routes, for example oral, subcutaneous, intramuscular, intravenous or intracerebral. The preferred route of administration would be oral at daily doses of about 0.01 -100 mg/kg.

This invention provides a method of treating metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases wherein the abnormality is improved by altering the activity of the cofactor thereby influencing the binding affinity of the cofactor to PXR, which could be useful for the treatment of disturbances in steroid homeostasis. Similarly, the invention also provides methods for treating diseases and conditions resulting from metabolic disorders, immunological indications, hormonal dysfunctions, neurosystemic diseases, or other diseases, which method comprises administering an effective amount of an agonist- or antagonist containing pharmaceutical composition described above.

TRANSFORMATION OF CELLS AND DRUG SCREENING :

The recombinant expression constructs of the present invention are useful in molecular biology to transform cells which do not ordinarily express the CFs to express these cofactors upon transformation.

Such cells are useful as intermediates for making cellular preparations useful for cofactor binding assays, which are in turn useful for drug screening.

The recombinant expression constructs of the present invention are also useful in gene therapy. Cloned genes of the present invention, or fragments thereof, may also be used in gene therapy carried out by homologous recombination or site-directed mutagenesis. See generally Thomas & Capecchi, Cell 51, 503-512 (1987); Bertling, Bioscience Reports 7, 107-112 (1987); Smithies et al., Nature 317, 230-234 (1985).

Oligonucleotides of the present invention are useful as diagnostic tools for probing cofactor gene expression in tissues. For example, tissues are probed *in situ* with oligonucleotide probes carrying detectable groups by conventional autoradiographic techniques, as explained in greater detail in the Examples below, to investigate

native expression of this cofactor or pathological conditions relating thereto. Further, chromosomes can be probed to investigate the presence or absence of the CF genes, and potential pathological conditions related thereto, as also illustrated by the Examples below. Probes according to the invention should generally be at least about 15 nucleotides in length to prevent binding to random sequences, but, under the appropriate circumstances may be smaller.

ANTIBODIES AGAINST THE COFACTOR PROTEIN OR POLYPEPTIDE

Another aspect of the invention includes antibodies specifically reactive with the proteins or any parts of the proteins according to the invention (SEQ ID NO. 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12 and/or SEQ ID NO. 30) and or polypeptides encoded by the nucleotide sequences of the cofactors or their complements (SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10, SEQ ID NO. 2, SEQ ID NO. 5, SEQ ID NO. 8, SEQ ID NO. 11, SEQ ID NO. 28 and/or SEQ ID NO. 29). (The term „antibody" refers to intact molecules as well as fragments thereof, such as Fa, F(ab).sub.2, and Fv, which are capable of binding the epitopic determinant.) By using immunogens derived from the polypeptide according to the invention and/or encoded by the nucleic acids according to the invention, anti-protein/anti-peptide antiserum or monoclonal antibodies can be made by standard protocols (E. Howell & D. Lane. Antibodies: A Laboratory Manual. *Cold Spring Harbor Laboratory* (1988)).

A polyclonal antibody is prepared by immunizing a mammal, such as a mouse, a hamster or rabbit with an immunogenic form of the cofactors according to the invention depending on which of these are desired) of the present invention, and collecting antisera from that immunized animal. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As an immunizing antigen, fusion proteins, intact polypeptides or fragments containing small peptides of interest can be used. They can be derived by expression from a cDNA transfected in a host cell with subsequent recovering of the protein/peptide or peptides can be synthesized chemically (*e.g.* oligopeptides with 10-15 residues in length). Important tools for monitoring the function of the cofactor

gene according to the present invention, *i.e.* encoded by a sequence according to SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and SEQ ID NO. 28 are antibodies against various domains of the proteins according to the invention.

A given polypeptide or polynucleotide may vary in its immunogenicity. It is often necessary to couple the immunogen (e.g. the polypeptide) with a carrier. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin (BSA) and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal in the presence of an adjuvant, a non-specific stimulator of the immune response in order to enhance immunogenicity. The production of polyclonal antibodies is monitored by detection of antibody titers in plasma or serum at various time points following immunization. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies. When a desired level of immunogenicity is obtained, the immunized animal may be bled and the serum isolated, stored and purified.

To produce monoclonal antibodies, antibody-producing cells (e.g. spleen cells) from an immunized animal (preferably mouse or rat) are fused by standard somatic cell fusion procedures with immortalizing cells such as myeloma cells to yield hybridoma cells. Where the immunized animal is a mouse, a preferred myeloma cell is the murine NS-1 myeloma cell. Such techniques are well known in the art, and include, for example, the hybridoma technique (originally developed by Kohler & Milstein. *Nature* 256: 495-497 (1975)), the human B cell hybridoma technique (Kozbar *et al. Immunology Today* 4:72 (1983)), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole *et al.* Monoclonal Antibodies and Cancer Therapy. *Alan R. Liss, Inc.* pp. 77-96 (1985)).

The fused spleen/myeloma cells are cultured in a selective medium to select fused spleen/myeloma cells from the parental cells. Fused cells are separated from the mixture of non-fused parental cells, for example, by the addition of agents that block the *de novo* synthesis of nucleotides in the tissue culture media. This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-clone dilution in microtiter plates, followed by testing the individual clonal supernatants for

reactivity with an antigen-polypeptide. The selected clones may then be propagated indefinitely to provide the monoclonal antibody in convenient quantity.

The creation of antibodies which specifically bind the polypeptides according to the invention and/or encoded by the nucleotide sequences of the cofactors or their complements provides an important utility in immunolocalization studies, and may play an important role in the diagnosis and treatment of such diseases and disorders as metabolic disorders, immunological indications, hormonal dysfunctions and/or neurosystemic diseases. The antibodies may be employed to identify tissues, organs, and cells which express the cofactors. Antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate cofactor protein levels in tissue or from cells in bodily fluid as part of a clinical testing procedure.

Monoclonal antibodies provided by the present invention are also produced by recombinant genetic methods well known to those of skill in the art, and the present invention encompasses antibodies made by such methods that are immunologically reactive with an epitope of a mammalian cofactor protein or peptide according to the invention.

The present invention encompasses fragments of the antibody that are immunologically reactive with an epitope of a cofactor protein or peptide. Such fragments are produced by any number of methods, including but not limited to proteolytic cleavage, chemical synthesis or preparation of such fragments by means of genetic engineering technology. The present invention also encompasses single-chain antibodies that are immunologically reactive with an epitope of a cofactor protein or peptide made by methods known to those of skill in the art.

CHIMERIC ANTIBODIES AND OTHER TYPES OF ANTIBODIES:

The invention also includes chimeric antibodies, comprised of light chain and heavy chain peptides immunologically reactive to an epitope that is a cofactor protein or peptide according to the invention. The chimeric antibodies embodied in the present invention include those that are derived from naturally occurring antibodies as well as

chimeric antibodies made by means of genetic engineering technology well known to those of skill in the art.

Also included are methods for the generation of antibodies against any of the group comprising the peptides according to SEQ ID NO: 3, SEQ ID NO. 6, SEQ ID NO. 9, SEQ ID NO. 12 and/or SEQ ID NO. 30 which rely on the use of phage display systems and related systems, such as described in Hoogenboom HR, de Bruine AP, Hufton SE, Hoet RM, Arends JW, Roovers RC, Immunotechnology 1998 Jun;4(1):1-20, and references therein.

EPITOPES OF THE COFACTORS

The present invention also encompasses one or more epitopes of a cofactor protein or peptide that is comprised of sequences and/or a conformation of sequences present in the cofactor proteins or peptide molecule. These epitopes may be naturally occurring, or may be the result of proteolytic cleavage of the cofactor proteins or peptides and isolation of an epitope-containing peptide or may be obtained by synthesis of an epitope-containing peptide using a method of genetic engineering technology and synthesized by genetically engineered prokaryotic or eukaryotic cells.

ANTISENSE OLIGONUCLEOTIDES AGAINST COFACTOR GENE TRANSCRIPTS

Antisense oligonucleotides are short single stranded DNA or RNA molecules which may be used to block the availability of the cofactor messenger(s). Synthetic derivatives of ribonucleotides or desoxyribonucleotides and/or PNAs (see above) are equally possible. These are potential candidate agents which may interact with the cofactor according to the invention.

The sequence of an antisense oligonucleotide is at least partially complementary to the sequence of the cofactor of interest. The complementarity of the sequence is in any case high enough to enable the antisense oligonucleotide to bind to the nucleic acid according to the invention or parts thereof (SEQ ID NO. 1, SEQ ID NO. 4, SEQ ID NO. 7, SEQ ID NO. 10 and/or SEQ ID NO. 28) in which the binding of

oligonucleotides to the target sequence interfere with the biological function of the targeted sequence (Brysch W, Schlingensiepen KH, Design and application of antisense oligonucleotides in cell culture, in vivo, and as therapeutic agents, Cell Mol Neurobiol 1994 Oct;14(5):557-68; Wagner RW, Gene inhibition using antisense oligodeoxynucleotides, Nature 1994 Nov 24;372(6504):333-5 or Brysch W, Magal E, Louis JC, Kunst M, Klinger I, Schlingensiepen R, Schlingensiepen KH Inhibition of p185c-erbB-2 proto-oncogene expression by antisense oligodeoxynucleotides down-regulates p185-associated tyrosine-kinase activity and strongly inhibits mammary tumor-cell proliferation, Cancer Gene Ther 1994 Jun;1(2):99-105 or Monia BP, Johnston JF, Ecker DJ, Zounes MA, Lima WF, Freier SM Selective inhibition of mutant Ha-ras mRNA expression by antisense oligonucleotides, J Biol Chem 1992 Oct 5;267(28):19954-62 or Bertram J, Palfner K, Killian M, Brysch W, Schlingensiepen KH, Hiddemann W, Kneba M, Reversal of multiple drug resistance in vitro by phosphorothioate oligonucleotides and ribozymes, Anticancer Drugs 1995 Feb;6(1):124-34)

This interference occurs in most instances at the level of translation, *i.e.* through the inhibition of the translational machinery by oligonucleotides that bind to mRNA, however, two other mechanisms of interference with a given gene's function by oligonucleotides can also be envisioned, (I) the functional interference with the transcription of a gene through formation of a triple helix at the level of genomic DNA and the interference of oligonucleotides with the function of RNA molecules that are executing at least part of their biological function in the untranslated form (Kochetkova M, Shannon MF, Triplex-forming oligonucleotides and their use in the analysis of gene transcription. Methods Mol Biol 2000;130:189-201 Rainer B. Lanz1, Nell J. McKenna1, Sergio A. Onate1, Urs Albrecht2, Jiemin Wong1, Sophia Y. Tsai1, Ming-Jer Tsai1, and Bert W. O'Malley A Steroid Receptor Coactivator, SRA, Functions as an RNA and Is Present in an SRC-1 Complex Cell, Vol. 97, 17–27, April, 1999).

Antisense oligonucleotides can be conjugated to different other molecules in order to deliver them to the cell or tissue expressing any of the cofactor genes. For instance the antisense oligonucleotide can be conjugated to a carrier protein (*e.g.* ferritin) in

order to direct the oligonucleotide towards the desired target tissue, *i.e.* in case of ferritin predominantly to the liver.

Antisense expression constructs are expression vector systems that allow the expression – either inducible or uninducible - of a complementary sequence to the cofactor sequences according to the invention. The potential possibility of such an approach has been demonstrated in many different model systems (von Ruden T, Gilboa E, Inhibition of human T-cell leukemia virus type I replication in primary human T cells that express antisense RNA, J Virol 1989 Feb;63(2):677-82; Nemir M, Bhattacharyya D, Li X, Singh K, Mukherjee AB, Mukherjee BB, Targeted inhibition of osteopontin expression in the mammary gland causes abnormal morphogenesis and lactation deficiency, J Biol Chem 2000 Jan 14;275(2):969-76; Ma L, Gauville C, Berthois Y, Millot G, Johnson GR, Calvo F Antisense expression for amphiregulin suppresses tumorigenicity of a transformed human breast epithelial cell line, Oncogene 1999 Nov 11;18(47):6513-20; Refolo LM, Eckman C, Prada CM, Yager D, Sambamurti K, Mehta N, Hardy J, Younkin SG, Antisense-induced reduction of presenilin 1 expression selectively increases the production of amyloid beta42 in transfected cells, J Neurochem 1999 Dec;73(6):2383-8; Buckley NJ, Abogadie FC, Brown DA, Dayrell M, Caulfield MP, Delmas P, Haley JE, Use of antisense expression plasmids to attenuate G-protein expression in primary neurons, Methods Enzymol 2000;314:136-48).

According to the invention an antisense expression construct can be constructed with virtually any expression vector capable of fulfilling at least the basic requirements known to those skilled in the art.

In one embodiment of the invention retroviral expression systems or tissue specific gene expression systems are preferred.

Current standard technologies for delivering antisense constructs are performed through a conjugation of constructs with liposomes and related, complex-forming compounds, which are transferred via electroporation techniques or via particle-mediated "gene gun" technologies into the cell. Other techniques may be envisioned by one skilled in the art.

Microinjection still plays a major role in most gene transfer techniques for the generation of germ-line mutants expressing foreign DNA (including antisense RNA constructs) and is preferred embodiment of the present invention.

RIBOZYMES DIRECTED AGAINST CF GENE TRANSCRIPT.

Ribozymes are either RNA molecules (Gibson SA, Pellenz C, Hutchison RE, Davey FR, Shillitoe EJ, Induction of apoptosis in oral cancer cells by an anti-bcl-2 ribozyme delivered by an adenovirus vector, Clin Cancer Res 2000 Jan;6(1):213-22; Folini M, Colella G, Villa R, Lualdi S, Daidone MG, Zaffaroni N, Inhibition of Telomerase Activity by a Hammerhead Ribozyme Targeting the RNA Component of Telomerase in Human Melanoma Cells, J Invest Dermatol 2000 Feb;114(2):259-267; Halatsch ME, Schmidt U, Botefur IC, Holland JF, Ohnuma T, Marked inhibition of glioblastoma target cell tumorigenicity in vitro by retrovirus-mediated transfer of a hairpin ribozyme against deletion-mutant epidermal growth factor receptor messenger RNA, J Neurosurg 2000 Feb;92(2):297-305; Ohmichi T, Kool ET, The virtues of self-binding: high sequence specificity for RNA cleavage by self-processed hammerhead ribozymes, Nucleic Acids Res 2000 Feb 1;28(3):776-783) or DNA molecules (Li J, Zheng W, Kwon AH, Lu Y, In vitro selection and characterization of a highly efficient Zn(II)-dependent RNA-cleaving deoxyribozyme; Nucleic Acids Res 2000 Jan 15;28(2):481-488) that have catalytic activity. The catalytic activity located in one part of the RNA (or DNA) molecule can be "targeted" to a specific sequence of interest by fusing the enzymatically active RNA molecule sequence with a short stretch of RNA (or DNA) sequence that is complementary to the cofactor gene transcript of interest. Such a construct will, when introduced into a cell either physically or via gene transfer of a ribozyme expression construct find the corresponding cofactor sequence (our sequence of interest or also targeted in RNA) and bind via its sequence-specific part to said sequence. The catalytic activity attached to the construct, usually associated with a special nucleic acid structure (people distinguish so called "hammerhead" structures and "hairpin" structures), will then cleave the targeted RNA. The targeted mRNA will be destroyed and cannot be translated efficiently, thus the protein encoded by the mRNA derived from cofactor will not be expressed or at least will be expressed at significantly reduced amounts.

These are potential candidate agents which may interact with the cofactor according to the invention.

In a preferred embodiment the invention covers inducible ribozyme constructs (Koizumi M, Soukup GA, Kerr JN, Breaker RR, Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP, Nat Struct Biol 1999 Nov;6(11):1062-1071).

In a further preferred embodiment the invention concerns the use of "bivalent" ribozymes (multimers of catalytically active nucleic acids) as described in (Hamada M, Kuwabara T, Warashina M, Nakayama A, Taira K, Specificity of novel allosterically trans- and cis-activated connected maxizymes that are designed to suppress BCR-ABL expression FEBS Lett 1999 Nov 12;461(1-2):77-85).

TRANSGENIC ANIMALS CARRYING THE CF1, CF2, CF3, CF4 AND/OR CF44 COFACTOR GENE

Also provided by the present invention are non-human transgenic animals grown from germ cells transformed with a CF1, CF2, CF3, CF4 or CF44 nucleic acid sequence according to the invention and that express the cofactor according to the invention and offspring and descendants thereof. Also provided are transgenic non-human mammals comprising a homologous recombination knockout of the native cofactors, as well as transgenic non-human mammals grown from germ cells transformed with nucleic acid antisense to the nucleic acids of the invention and offspring and descendants thereof. Further included as part of the present invention are non-human transgenic animals in which the native cofactor has been replaced with the human ortholog. Of course, offspring and descendants of all of the foregoing transgenic animals are also encompassed by the invention.

Transgenic animals according to the invention can be made using well known techniques with the nucleic acids disclosed herein. E.g., Leder et al., U.S. Patent Nos.4,736,866 and 5,175,383; Hogan et al., Manipulating the Mouse Embryo, A Laboratory Manual (Cold Spring Harbor Laboratory (1986)); Capecchi, Scienc 244,

1288 (1989); Zimmer and Gruss, Nature 338, 150 (1989); Kuhn et al., Science 269, 1427 (1995); Katsuki et al., Science 241, 593 (1988); Hasty et al., Nature 350, 243 (1991); Stacey et al., Mol. Cell Biol. 14, 1009 (1994); Hanks et al., Science 269, 679 (1995); and Marx, Science 269, 636 (1995). Such transgenic animals are useful for screening for and determining the physiological effects of the cofactor agonists and antagonist.

Consequently, such transgenic animals are useful for developing drugs to regulate physiological activities in which the cofactors participate.

The following Examples are provided for illustrative purposes only and are not intended, nor should they be construed, as limiting the invention in any manner.

MODELLING OF THE STRUCTURE OF CF1, CF2, CF3, CF4 AND/OR CF44

In one embodiment of the invention the amino acid sequences of the present invention can be used for structural drug design. Aim is to produce structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g. agonists, antagonists or inhibitors) in order to design drugs which are, for example, more active or stable forms of the polypeptide, or which, for example, enhance or interfere with the function of a polypeptide *in vivo*. In one approach one first determines the three-dimensional structure of a protein of interest, *i.e.* the cofactor, by computer-modeling, x-ray crystallography or a combination of both approaches. Additional useful information with respect to the structure of a polypeptide could also be gained from comparison of the protein sequence of the protein of interest with the sequence of related proteins where the structure is known. From the three-dimensional structure, binding sites of potential inhibitors or activators can be predicted. It can further be predicted which kinds of molecule might bind there. The predicted substances can then be screened to test their effect on the activity of the protein and its biological function.

EXAMPLES

EXAMPLE 1: CLONING AND EXPRESSION OF THE GENES ACCORDING TO THE INVENTION

Construction of suitable vectors containing the desired coding and control sequences employs standard ligation and restriction techniques that are well understood in the art. Isolated plasmids, DNA sequences, or synthesized oligonucleotides are cleaved, tailored, and religated in the form desired.

Site-specific DNA cleavage is performed by treatment with the suitable restriction enzyme (or enzymes) under conditions that are generally understood in the art, and the particulars of which are specified by the manufacturer of these commercially available restriction enzymes.

See, e.g., New England Biolabs, Product Catalog. In general, about 1 µg of plasmid and/or DNA sequence is cleaved by one unit of enzyme in about 20 µl of buffer solution. Often excess of restriction enzyme is used to ensure complete digestion of the DNA substrate. Incubation times of about one hour to two hours at about 37°C are workable, although variations are tolerable.

After each incubation, protein is removed by extraction with phenol/chloroform, and may be followed by ether extraction. The nucleic acid may be recovered from aqueous fractions by precipitation with ethanol. If desired, size separation of the cleaved fragments may be performed by polyacrylamide gel or agarose gel electrophoresis using standard techniques. A general description of size separations is found in Methods in Enzymology 65, 499-560 (1980).

Transformed host cells are cells which have been transformed or transfected with recombinant expression constructs made using recombinant DNA techniques and comprising cofactor encoding sequences. Preferred host cells for transient transfection are COS-7 cells. Transformed host cells may ordinarily express one of the cofactors CF1, CF2, CF3, CF4 or CF44, but host cells transformed for purposes of cloning or amplifying nucleic acid hybridization probe DNA need not express the cofactors. When expressed, the cofactor proteins will typically be located in the host cell membrane.

Cultures of cells derived from multicellular organisms are desirable hosts for recombinant nuclear receptor protein synthesis. In principal, any higher eukaryotic cell culture is workable, whether from vertebrate or invertebrate culture. However, mammalian cells are preferred. Propagation of such cells in cell culture has become a routine procedure. See Tissue Culture (Academic Press, Kruse & Patterson, Eds., 1973). Examples of useful host cell lines are bacteria cells, insect cells, yeast cells, human 293 cells, VERO and HeLa cells, LMTK- cells, and WI138, BHK, COS-7, CV, and MDCK cell lines. Human 293 cells are preferred.

EXAMPLE 2: COFACTOR CF1, CF2, CF3, CF4 AND CF44 OR TISSUE LOCALIZATION:

A multiple tissue northern blot (Clontech, Palo Alto) is hybridized to a labeled probe. The blot contains about 0.3 to 3 µg of poly A RNA derived from various tissues. Hybridization may be carried out in a hybridization solution such as one containing SSC (see Maniatis et al, ibid) at an optimized temperature between 50°c and 70°C, preferably 65°C. The filter may be washed and a film exposed for signal detection (see also: Maniatis et al., *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y.(1989)).

EXAMPLE 3: COFACTOR cDNA ISOLATION FROM HUMAN AND OTHER ORGANISMS:

A cloning strategy is used to clone the CF1, CF2, CF3, CF4 or CF44 cofactor cDNA from specific cDNA libraries (Clontech, Palo Alto) or alternatively, RNA is obtained from various tissues and used to prepare cDNA expression libraries by using for example an Invitrogen kit. (Invitrogen Corporation, San Diego). For the isolation of the CF cDNA clones the chosen library may be screened under stringent condition (see definitions above) by using CF1, CF2, CF3, CF4 or CF44 specific probes. The cDNA insert of positive clones is subsequently sequenced and cloned in a suitable expression vector.

Additionally, full length cofactor clones from various species are obtained by using RACE PCR technology. In brief, suitable cDNA libraries are constructed or purchased. Following reverse transcription, the first strand cDNA is used directly in RACE PCR reactions using a RACE cDNA amplification kit according to the manufactures protocol (Clontech, Palo Alto). Amplified fragments are purified, cloned and subsequently used for sequence analysis.

To obtain information about the genomic organization of the cofactor gene, genomic libraries (Clontech, Palo Alto) are screened with a receptor specific probe under stringent conditions. Positive clones are isolated and the complete DNA sequence of the putative receptor is determined by sequence analysis (Maniatis et al., *Molecular Cloning: A laboratory Manual*, Cold Spring Harbor Laboratory Press, N.Y.(1989)).

EXAMPLE 4: ISOLATION OF THE COFACTOR PROTEINS BY USE OF THE YEAST TWO-HYBRID SYSTEM

A yeast two-hybrid assay was performed using methods such as described by Fields and Song *Nature* 340, pp245 (1989), Bartel et al., *Biotechniques* 14, pp920 (1993) and Lee et al. Nature 374 pp91-4 (1995). A sequence encoding amino acids 106-434 of PXR (containing the ligand binding domain; LBD) was cloned into the vector pGBT9 (Clontech) in such way that, after transformation of the haploid yeast strain CG1945 (Clontech), a hybrid protein is expressed consisting of the DNA-binding domain (BD) of the Gal4 transcription factor fused N-terminally to amino acids 106-434 of PXR. CG1945 cells expressing the Gal4BD::PXR fusion protein were mated to cells of strain Y187 (Clontech) containing a library of Gal4 transcription activation domain (AD) fusion plasmids with human cDNA generated from a range of tissues inserted into the vector pACT2 (Clontech). All libraries were purchased from Clontech Laboratories (MATCHMAKER human cDNA libraries) and included Cat. numbers HL4040AH (aorta), HL4041AH (chondrocytes), HY4004AH (brain), HY4035AH (testis), HY4024AH (liver), HY4042AH (heart), HY4053AH (bone marrow) and HY4028AH (fetal brain). The two-hybrid screens were essentially performed following the Clontech "Pretransformed Matchmaker Libraries User Manual" (PT3183-1): Transformed CG1945 and Y187 cells were mated in order to coexpress the Gal4::PXR fusion protein and the Gal4AD fusion proteins encoded on the library

plasmids within one cell. Interaction of the two hybrid proteins led to activation of reporter gene transcription. Cells were selected for interactions of PXR with library proteins on medium lacking tryptophan, leucine and histidine and were further assayed for expression of α-galactosidase, encoded by the *MEL1* reporter gene. Colonies which were positive for reporter gene activation were chosen for further analysis. The DNA inserts of the library plasmids contained in these colonies were amplified by use of the polymerase chain reaction directly on the yeast colonies using oligonucleotide primer which hybridize on vector sequences flanking both sides of the insert. The identity of the insert was determined by standard DNA sequencing techniques.

Five novel cofactors interacting with PXR were isolated using this approach: CF1 was isolated from the heart cDNA library, CF2 from the aorta, testis, fetal brain and brain cDNA libraries, CF3 from the fetal brain and heart cDNA libraries, CF4 from the heart cDNA library and CF44 from the chondrocyte library.

EXAMPLE5: DETECTION OF MUTANT ALLELES OF THE GENE(S) ACCORDING TO THE INVENTION AND THEIR UTILISATION FOR DIAGNOSTIC PURPOSES.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type cofactor gene is detected. In addition, the method can be performed by detecting the wild-type cofactor gene and confirming the lack of cause of the disease as a result of the locus.

"Alteration of the wild-type gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and non-coding regions. Deletions may be of the entire gene or of only a portion. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions. Somatic mutations are those which occur only in certain tissues and are not inherited in the germline. Germline mutations can be found in any of a body's tissue and are mostly inherited. Point mutational events may occur in regulatory regions, such as the promotor of the gene, leading to loss or dimunition of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the cofactor gene product or to a decrease in mRNA stability or translation efficiency.

Applicable diagnostic techniques include, but are not limited to fluorescent in situ hybridization (FISH), direct DNA sequencing, PFGE analysis, Southern blot analysis, single stranded conformation analysis (SSCA), RNAse protection assay, allele-specific oligonucleotide (ASO), dot blot analysis, hybridization using nucleic acid modified with gold nanoparticles and PCR-SSCP, as discussed in detail further below. Furthermore, DNA microchip technology can be applied.

The presence of a disease due to a germline mutation of a cofactor can be ascertained by testing any tissue of the diseased human for mutations of the cofactor gene. For instance, a person who has inherited a germline mutation in the cofactor gene, especially one that will alter the interaction of the cofactor with the PXR protein, will be prone to develop a disease. The presence of such a mutation can be determined by extracting DNA from any tissue of the body. For example, blood can be drawn and DNA extracted from blood cells and analyzed. Moreover, prenatal diagnosis of the disease will be possible by testing fetal cells, placental cells or amniotic cells for mutations in the cofactor gene. There are several methods that allow the detection of alterations of the wild-typ cofactor gene, including for instance point mutations as well as deletions in the DNA sequence and these methods are discussed here:

Direct genomic DNA Sequencing, either manual or by automated means can detect sequence variations of cofactor genes (Nucleic Acids Res 1997 May 15;25(10):2032-2034 Direct DNA sequence determination from total genomic DNA. Kilger C, Pääbo S, Biol. Chem. 1997 Feb; 378(2):99-105, Direct exponential amplification and sequencing (DEXAS) of genomic DNA. Kilger C, Pääbo S, DE 19653439.9 and DE 19653494.1). Another way is to make use of the single-stranded conformation polymorphism assay (SSCP; Orita et al., PNAS 86, 2766 (1989)). Variations in the DNA sequence of the cofactor gene from the wild-type sequence will be detected due to a shifted mobility of the corresponding DNA-fragments in SSCP gels.

Other approaches are based on the detection of mismatches between the two complementary DNA strands. These methods, which will not allow the detection of large deletions, duplications or insertions nor the detection of a regulatory mutation

affecting transcription or translation of the cofactor gene include the clamped denaturing gel electrophoresis (CDGE; Sheffield et al.,1991), heteroduplex analysis (HA; White et al., *Genomics* 4, 560 (1992)) and chemical mismatch cleavage (CMC; Grompe et al., 1989). Other methods detect specific types of mutations such as deletions, duplications or insertions, for instance a protein truncation assay or the asymmetric assay. These assay however, will not detect missense mutations. A review of currently available methods of detecting DNA sequence variation can be found in a review by Grompe, *Nature Genetics* 5, 111 (1993). Once a mutation is known, an allele specific detection approach such as allele specific oligonucleotide (ASO) hybridisation will allow the rapid screening of a large number of other sample for that mutation. Such a technique may involve the utilisation of probes which are labeled with gold nanoparticles to to yield a visual colour result (Elghanian et al.,Science 277, 1078 (1997)).

In another embodiment of the present invention large scale genetic studies might be applied to investigate the association of a disease-phenotype with the gene of interest. The availability of the human genome allows an easy definition of genetic markers for most genes for a particular disease physiology. More importantly, single nucleotide polymorphisms (SNPs) are amenable markers for large genetic studies. SNPs in coding or regulatory regions of genes which are thought to contribute to a disease physiology can have a direct impact on the phenotype, *e.g.* change a quantitative readout of disease physiology, for example the age of onset of heart attack. Association and linkage studies with related individuals, therefore provide an excellent means to test or verify a hypothesis on the functional impact of the gene of interest on disease physiology in vivo, in humans.

The PXR protein is known to be involved in controlling the expression of the cyclooxgenase P450 or Cyp3A gene. The product of the Cyp3A gene product is a hydroxylase which is involved in the metabolism of xenobiotics, including the majority of drugs in use. Therefore, alterations identified in the PXR gene can be used to predict the ability of a given human individual to metabolize xenobiotics. Proteins interacting with PXR, such as the cofactors according to the invention will also be involved in the function of PXR. Therefore, alterations in the cofactors are useful for

determining the genetic state of a person with respect to its abilitiy to metabolise xenobiotics, including drugs.

One embodiment of the invention is the use of genetic testing of individuals for alterations in the genes of the cofactor which interact with PXR to try and predict the individuals' capabilities of metabolising xenobiotic substances. This can be done using any of the methods described to determine which alleles are present in the genome of a given person. In particular, one embodiment is the use of the information on polymorphisms in the genes coding for the cofactors to try and predict the metabolism of certain drugs. Such drugs include substances which are metabolized via the product of the Cyp3a4, or related enzymes. In one preferred embodiment, the information on the genetic variations in the cofactors genes is used to predict the effect of hyperforin (St Johns worth) on a individual. In another preferred embodiment, the information on the genetic variations in the cofactor genes is used to predict the effect of steroids and steroid derivatives, such as 17beta estradiol and dexamethasone or pregnenolone. The information of the genetic state of an individual is also useful to predict drug-drug interactions.

In order to detect polymorphisms in DNA sequences, DNA samples can be prepared from normal individuals and from persons being affected by the disease and these samples can be cut by one or more restriction enzymes and applied to Southern analysis. Southern blots displaying hybridizing fragments differing in length from the control DNA when probed with sequences near or including the cofactor locus could indicate a possible mutation. If large DNA fragments are used it is appropriate to separate these fragments by pulsed field gel electrophoresis (PFGE).

Detection of point mutations may be accomplished by amplification, for instance by PCR, from genomic or cDNA and sequencing of the amplified nucleic or by molecular cloning of the cofactor allele and sequencing the allele using techniques well known in the art.

There are six well known methods for a more complete, yet still indirect, test for confirming the presence of a susceptibility allele: 1) single stranded conformation analysis (SSCP) (Orita et al., *PNAS*, 86, 2766 (1989)); 2) denaturing gradient gel electrophoresis (DGGE) (Wartell et al., *NAR* 18, 2699, (1990); Sheffield et al., *PNAS*

86 , 232 (1989)); 3) RNase protection assays (Finkelstein et al., *Genomics* 7, 167 (1990); Kinszler et al., Science 251, 1366 (1991)); 4) allele specific oligonucleotides (ASOs, Conner et al., PNAS, 80, 278 (1983)); 5) the use of proteins which recognise nucleotide mismatches, such as the E. coli mutS protein (Modrich, *Ann. Rev. Genetics*, 25, 229 (1991)) and 6) allele-specific PCR (Ruano and Kidd, *NAR* 17, 8392 (1989)). For allele-specific PCR, primers are used which hybridise at their 3' ends to a particular *cofacto* mutation. Without the mutation, no PCR product is observed. Amplification Refractory Mutation System could also be used, as disclosed in European Patent Application Publication No. 0332435 and in Newton et al., *NAR* 17, 2503 (1989). Insertions and deletions of genes can also be detected by molecular cloning, amplification and sequencing. Moreover, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to score for alteration of an allele or an insertion in a polymorphic fragment. Such a method would be particularly useful for screening relatives of an affected person for the presence of the mutation found in that person. Other approaches for detecting insertions and deletions as known for those trained in the art can be used.

SSCP detects a band which migrates differently because the variation causes a difference in single strand, intra molecular base pairing. The RNAse protection assay involves cleavage of the mutant fragment into two or more smaller fragments. By using DGGE variations in the DNA can be detected by differences in the migration rates of mutant compared to normal alleles in a denaturing gradient gel. In the mutS assay, the protein binds only to sequences that contain a nucleotide mismatch in a hetero duplex between mutant and wild-type sequences.

Mismatches, according to the present invention, are hybridised nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or the corresponding mRNA product. While these techniques are less sensitive than sequencing, they can preferably be used when a large number of samples shall be tested. An example of the a mismatch cleavage method is the RNAse protection assay. In the practice of the present invention , the method involves the use of a labeled ribonucleotide probe which is complementary to the wild-type sequence of the cofactor gene coding sequence.The

riboprobe and either mRNA or DNA isolated from the person are hybridised together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by the enzyme, it cleaves at the site of the mismatch. Consequently, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNAse A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA. If the riboprobe comprises only a fragment of the mRNA or the gene, it is advantageous to use a number of probes to screen the whole mRNA sequence for mismatches.

Similarly, DNA probes can be used to detect mismatch mutations through enzymatic or chemical cleavage (Cotton et al., *PNAS* 85, 4397 (1988); Shenk et al., *PNAS* 72, 989 (1975); Novack et al., *PNAS* 83, 586 (1986)). Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to match duplexes (Cariello, *Human Genetics* 42, 726 (1988)). With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridisation. Variations in DNA of the *cofactor)* gene can also be detected using Southern hybridisation, especially if the changes are major rearrangements, such as deletions or insertions. DNA sequences of the cofactor gene which have been amplified by PCR may also be screened using allele specific probes. These probes are nucleic acid oligomers, each of which contains a region of the gene sequence harboring a known mutation. For instance, one oligomer could be about 25 nucleotides in length corresponding to a portion of the gene sequence. By using a number of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously discovered mutation in the gene. Hybridisation of allele-specific probes with amplified *cofactor* sequences can be performed, for example, on a nylon filter. Under high stringency hybridisation conditions, the hybridisation of a particular probe should indicate the presence of the same mutation in the tissue as in the allele-specific probe.

The newly developed technique of nucleic acid analysis via microchip technology is also applicable to the present invention. In this technique, thousands of distinct nucleotide probes are built up in an array on a silicon chip. Nucleic acid to be

analysed is fluorescently labeled and hybridised to the probes on the chip. It is also possible to study nucleic acid-protein interactions using these nucleic acid microchips. Using this technique one can determine the presence of mutations or even sequence the nucleic acid being analysed or one can measure expression of a gene of interest. This method is one of parallel processing of thousands of probes at once and can tremendously accelerate the analysis. In several publications the use of this method is described (Hacia et al., *Nature Genetics* 14, 441 (1996); Shoemaker et al., *Nature Genetics* 14, 450 (1996); Chee et al., *Science* 274, 610 (1996); DeRisi et al., *Nature Genetics* 14, 457 (1996)). This new technology has also been reviewed in Borman et al., *Chemical and Engineering News* 9, 42 (1996) and has been subject of an editorial in *Nature Genetics* (1996).

The most definite test for mutations in a candidate locus is to directly compare genomic *cofactor* sequences from patients with those from normal individuals. Alternatively one could sequence mRNA after amplification (for example by PCR) thereby eliminating the necessity of determining the exon structure of the respective gene.

Mutations from patients falling outside the coding region of the cofactor gene can be detected by examining the noncoding regions, such as introns and regulatory sequences within or near the genes. Early indications of mutations in noncoding regions could be for example the abundance or abnormal size of mRNA products in patients as compared to control individuals as detected by northern blot analysis.

Alteration of cofactor expression can be detected by any techniques known in the art. These include northern blot analysis, PCR amplification and RNAse protection. Diminished mRNA expression indicates an alteration in the wild-type gene sequence. Alterations of wild-type genes can also be detected by screening for alteration of cofactor protein. For example, monoclonal antibodies against cofactor protein can be used to screen a tissue. Lack of cognate antigen would indicate a mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant gene product. These kind of immunological assays could be done in any convenient format known in the art. These include western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered cofactor protein can be used

to detect alteration of the wild-type *cofactor* gene. Functional assays such as protein binding determinations can be used. Moreover, assays can be used which detect the cofactor's biochemical function. Finding a mutant cofactor gene product indicates an alteration of the cofactor wild-type gene. One such binding assay would test the binding of cofactor protein with wild-type PXR protein. Conversely, wild-type PXR protein or the domain interacting with the cofactor protein can be used in a protein binding assay or biochemical function assay to detect normal or mutant PXR proteins.

A mutant cofactor gene or gene product or a mutant PXR protein can also be detected in other human body samples, such as serum, stool, urine and sputum. The same techniques discussed above for detection of mutant genes or gene products in tissues can be applied to other body samples. By screening such body samples, a simple early diagnosis can be achieved for the disease) resulting from a mutation in the cofactor gene.

EXAMPLE 6: A CELL BASED ASSAY FOR MEASURING THE BINDING OF THE COFACTOR TO PXR.

The DNA sequence encoding the open reading frame of the cofactor is transferred into the vector pVP16 (Clontech) to allow the expression of a fusion protein of the cofactor with the strong transactivation domain of the VP16 protein (of herpex simplex virus) in mammalian cells under the control of the strong CMV promoter. On another vector (the reporter), the luciferase gene is cloned under the control of a minimal promoter containing a PXR-responsive DNA element. This vector also expresses a second enzyme, *e.g.* beta-galactosidase, under the control of a constitutive promoter, to allow normalization for transfection efficiency between experiments. A third vector contains the PXR gene under the control of the strong CMV promoter.

CV-1 cells are then transiently transfected with different combinations of the three plasmids. Transfection is done by standard methods, e.g. by use of the CalPhos Maximizer (Clontech, #8021-1,-2). Interaction of the cofactor protein with PXR will lead to a strong transactivation due to the attached VP16 domain of the cofactor

fusion protein. Thus, interaction of the cofactor with PXR will result in increased luciferase activity. Thus, inclusion of the cofactor VP16 will result in increased luciferase activity as compared to transfection of the PXR and the reporter alone. To measure this effect, extracts are prepared of the transfected cells 48 to 72 hours after transfection, and luciferase activity is determined. To normalize for transfection efficiency, beta-galactosidase activity is also determined.

Addition of substances known or suspected to influence the binding of PXR to the cofactor are added to the medium of the transfected cells. These substances are added at different time points prior to cell lysis, typically ranging between 18 hours to a five minutes before cell lysis. Luciferase activity is taken as a measure of the effect of these substances on the binding of the cofactor to PXR. To avoid activation of PXR by substances contained in the serum of the medium, charcoal-stripped serum has to be used for these experiments.

In an alternative setting of the experiment, the DNA-binding domain of PXR is replaced with the DNA-binding domain of the yeast GAL4 transcription factor. On the reporter plasmid, the luciferase is expressed under the control of GAL4-responsive upstream acitivating sequences. Expression of luciferase again is an indication for binding of the cofactor-VP16 fusion to the PXR-GAL4 fusion. This setting is also referred to as the mammalian two hybrid system. A description of the experiment is found in the manual to the Mammalian MATCHMAKER Two-Hybrid Assay Kit from Clontech, # PT3002-1, catalogue #K1602-1)

Substances activating nuclear receptors cause an exchange of the proteins bound to the receptors, thus effecting the dissociation of some proteins and promoting the binding of other proteins. Thus, in the experiments as described above, one can test for PXR-activating compounds and PXR-inactivating compounds by monitoring the binding of the cofactor to PXR.

In an alternative setting, stably transfected cell lines are used which contain copies of the two different expression constructs for PXR and the cofactor as well as the reporter construct stably integrated into the chromosomes of the cells.

EXAMPLE5: A FRET ASSAY USING COFACTOR PROTEINS

DNA sequences encoding the open reading frame of the cofactor and the PXR gene are each transferred separately into the vector pENTRY (Life Technologies) to allow efficient construction of a diverse set of expression constructs. The open reading frame is then recombined into the vector pDEST17 for expression in E. coli strain BL21 as a fusion protein to a six-histidine tag induced by IPTG, as well as into the pDEST15 for expression as a fusion protein with glutathione S-transferase (GST). The plasmids pDEST15, pDEST17and pENTRY are purchased from LIFE TECHNOLOGIES. Alternatively, the open reading frame is introduced into the vector pLV-CBDgw for expression as a fusion protein with the calmodulin binding protein using recombinant baculoviruses as specified by the manufacturer (Becton Dickinson). pLV-CBDgw is a derivative of the vector pLV1392 (Becton Dickinson) which is modified by the insertion of a calmodulin binding protein fragment, followed by the sequence required for recombinational cloning via the Gateway system (Life Technologies). Protein expression is induced and recombinant protein is purified by passage over a Ni-NTA-column, or a glutathione column or a calmodulin column, respectively.

To measure the interaction of the two proteins, a biotinylated (Biotintag Micro biotinylation Kit, Sigma) His-tagged PXR protein and the GST fusion of the cofactor are mixed at 0.2-5 µM. Antibody to the GST protein is added which is labelled by the europium chelate at a concentration of 1-3 (typical 2.5) nM. Streptavidin which is fluorescently labeled by covalent attachment of allophycocyanin is added at a concentration of 5-30 µg/ml (typical 10µg/ml). The europium chelate is stimulated by a flash of light (320nm) and, the emitted light is measured in a delayed (50-200 µs) time window for 300 to 600 µs after the flash at 615 nm (fluorescence of europium chelate) and 655nm (fluorescence of APC). Since APC is only excited by the light emitted by the europium chelate, a close proximity of the two different fluorophores is required for excitation. The strength of the APC signal, as well as the ratio of the signals from the two fluorophores (i.e. the ratio of the intensities of light emitted at 655 and 615nm) serves as a measure for the interaction of the two proteins.
R action buffers contain 20mM TrisHCl pH 7.9, 60mM KCl, 4mM $MgCl_2$. Reaction volume is 25µl. The Wallac VictorV fluorimeter is used for the fluorimetric measurements.

In an alternative setting, the cofactor is used as a biotinylated His-tagged protein, and the PXR protein is used as fusion to GST. In yet another setting, the His-tagged proteins are replaced by the same proteins fused to the calmodulin binding protein. In the latter case, the detection of the interaction is via biotinylated calmodulin, which is in turn binding to APC-coupled streptavidin. Calcium has to be included in the buffer in the form of 4mM $CaCl_2$, to allow complex formation between calmodulin and the calmodulin binding protein.

FIGURE CAPTIONS:

Fig. 1 shows sequences from the cofactor CF1 according to the invention, *i.e* cDNA sequence, reverse complement of the cDNA sequence and protein sequence. .

Fig. 2 shows sequences from the cofactor CF2 according to the invention *i.e* cDNA sequence, reverse complement of the cDNA sequence and protein sequence..

Fig. 3 shows sequences from the cofactor CF3 according to the invention *i.e* cDNA sequence, reverse complement of the cDNA sequence and protein sequence..

Fig. 4 shows sequences from the cofactor CF4 according to the invention *i.e* cDNA sequence, reverse complement of the cDNA sequence and protein sequence..

Fig. 5 shows sequences from PXR.

Fig. 6 shows the Ligand Binding Domain from PXR.

Fig. 7 shows sequences from RXR alpha.

Fig. 8 shows sequences from RXR beta.

Fig. 9 shows sequences from RXR gamma.

Fig. 10 shows sequences from the cofactor CF44 according to the invention *i.e* cDNA sequence, reverse complement of the cDNA sequence and protein sequence..

CLAIMS:

1. An isolated nucleic acid molecule coding for a cofactor of the pregnane x nuclear receptor which is selected from the group comprising:

   a) the nucleotide sequences set forth in SEQ ID NOs: 1 and 28;

   b) or complements thereof as set forth in SEQ ID NOs: 2 and 29;

   c) a nucleic acid which hybridizes to a nucleic acid having a nucleotide sequence which is the complement of the nucleotide sequence of SEQ ID NOs 1 and 28 under conditions of high stringency, and

   d) a nucleic acid which hybridizes to a nucleic acid having a nucleotide sequence which is the complement of the nucleotide sequence of SEQ ID NOs: 2 and 29 under conditions of high stringency.

2. The isolated nucleic acid molecule of claim 1 which is genomic DNA.

3. The isolated nucleic acid molecule of claim 1 which is cDNA.

4. The isolated nucleic acid molecule of claim 1 which is RNA.

5. An isolated nucleic acid molecule comprising the nucleic acid molecule of any of claims 1 to 4 and a label attached thereto.

6. A vector comprising the nucleic acid molecule of claim 1.

7. The vector of claim 6, which is an expression vector.

8. A host cell transfected with the vector of claim 6 or 7.

9. A host cell transfected with the expression vector of claim 7.

10. A method of producing a polypeptide comprising the step of culturing the host cell of claim 9 in an appropriate culture medium to, thereby, produce the polypeptide.

11. An isolated polypeptide encoded by any portion of the nucleic acid of claim 1.

12. An isolated polypeptide selected from the group comprising:
the amino acid sequences set forth in SEQ ID NOs.: 3 and/or 30.

13. Complex comprising a cofactor polypeptide according to any of SEQ ID NOs. 3 or 30 or a portion thereof additionally,
comprising a PXR polypeptide according to any of SEQ ID NOs. 15 or 18 or a portion thereof.

14. Complex comprising a cofactor polypeptide according to any of SEQ ID NOs. 3 or 30 or a portion thereof additionally,
comprising a PXR polypeptide according to any of SEQ ID NOs. 15 or 18 or a portion thereof additionally,
comprising a RXR polypeptide according to any of SEQ ID NOs. 21, 24 or 27 or a portion thereof.

15. Complex comprising a cofactor polypeptide according to any of SEQ ID NOs. 3 or 30 or a portion thereof additionally,
comprising a RXR polypeptide according to any of SEQ ID NOs. 21, 24 or 27 or a portion thereof.

16. A method for screening for agents which are capable of inhibiting the cellular function of the cofactor CF1 and/or CF44, comprising the steps of:

a) contacting one or more candidate agents with a polypeptide according to claims 11, 12 or a complex according to claims 13, 14, or 15,
b) removing unbound agent(s)
c) detecting whether the agent(s) interact with the polypeptide of the cofactor.

17. A method for screening for agents which are capable of inhibiting or activating the cellular function of PXR, comprising the steps of:

a) contacting one or more candidate agents which are capable of binding a complex according to claims 13, 14, or 15, with a complex according to claims, 13, 14 or 15,
b) removing unbound agent(s),
c) detecting the amount of the polypeptide according to any of claims 11 or 12 of the cofactor that has remained bound within the complex and
d) identifying such agents capable of either: i) releasing a large amount the polypeptide according to any of claims 11 or 12 of the cofactor from the complex, or ii) promoting the association of polypeptides according to any of claims 11 or 12 of the cofactor to the complex.

18. Agent identified by the method according to claim 17.

19. A method for inhibiting or activating the cellular function of the cofactor CF1, and/or CF44, comprising the steps of:

a) contacting a cell with a binding agent that binds the polypeptide according to claim 11, 12 or the complex according to claims 13,14 or 15,
b) whereby the cellular function of CF1, CF2, CF3 or CF4 is inhibited or activated.

20. A method for inhibiting or activating the binding of the cofactor CF1 and/or CF44, to a PXR polypeptide according to any of SEQ ID NOs. 15 or 18 or a portion thereof comprising the steps of:

a) contacting the polypeptide according to claim 11, 12 or the complex according to claims 13,14 or 15 with a binding agent,
b) whereby the binding of CF1 or CF44 to the PXR polypeptide is inhibited or activated.

21. Method according to claim 19 or 20,
characterized in that the binding agent is an antibody.

22. Method according to claim 19 or 20,
characterized in that the binding agent is RNA.

23. Method according to claim 19 or 20,
characterized in that the binding agent is an anti-sense oligonucleotide.

24. Method according to claim 19 or 20,
characterized in that the binding agent is a ribozyme.

25. Method according to claim 19 or 20,
characterized in that the binding agent is a steroid molecule.

26. Method according to claim 19 or 20,
characterized in that the cell is in a body.

27. A method for predicting the ability of a human being to metabolise xenobiotic substances or drugs comprising the steps of:
a) screening the genes coding for CF1 and CF44 for genetic aberrations or polymorphisms,
b) whereby the existence of a genetic aberration or polymorphism will predict an altered ability to metabolise xenobiotic substances or drugs.

28. Use of the proteins or a portion thereof according to SEQ ID NO. 3 and/or SEQ ID NO. 30 or a complex according to claims 13, 14 or 15 for the screening for substances that bind said proteins or portions thereof or complexes.

29. Use according to claim 28 wherein the screening is for agonists or antagonist of PXR.

Fig. 1

CF1:

SEQ ID NO. 1.:

CTTGCTCTGGCTGTTTCTGCCCCTGGGTTAACATTCAAGATGGTACATGCTGA
AGCCTTTTCTCGTCCTTTGAGTCGGAATGAAGTTGTTGGTTTAATTTTCCGTT
TGACAATATTTGGTGCAGTGACATACTTTACTATCAAATGGATGGTAGATGCA
ATTGATCCAACCAGAAAGCAAAAAGTAGAAGCTCAGAAACAGGCAGAAAAACT
AATGAAGCAAATTGGAGTGAAAAATGTGAAGCTCTCAAAATATGAAATGAGTA
TTGCTGCTCATCTTGTAGACCCTCTTAATATGCATGTTACTTGGAGTGATATA
GCAGGTTTAGATGATGTCATTACGGATCTGAAAGACACAGTCATCTTACCTAT
CAAAAAGAAACATTTGTTTGAGAATTCCAGGCTTCTGCAGCCTCCAAAAGGTG
TTCTTCTCTATGGGCCTCCAGGCTGGGGTAAAACGTTGATTGCCAAGGCCACA
GCCAAAGAAGCAGGCTGTCCATTTATTAACCTTCAGCCTTCGACACTGACCGA
TAAGTGGTATGGAGAATCTCAGAAATTGGCTGCTGCTGTCTTATCCCTTGCCA
TAAAGCTACAACCATCCATCATCTTTATAGATGGAAATAGACTCCTTTTTTAC
GAAACCGTTCAAGTTCTGACCATGAAAGCTACAGCCCAT

SEQ ID NO. 2:

REVERSE COMPLEMENT

ATGGGCTGTAGCTTTCATGGTCAGAACTTGAACGGTTTCGTAAAAAAGGAGTC
TATTTCCATCTATAAAGATGATGGATGGTTGTAGCTTTATGGCAAGGGATAAG
ACAGCAGCAGCCAATTTCTGAGATTCTCCATACCACTTATCGGTCAGTGTCGA
AGGCTGAAGGTTAATAAATGGACAGCCTGCTTCTTTGGCTGTGGCCTTGGCAA
TCAACGTTTTACCCCAGCCTGGAGGCCCATAGAGAAGAACACCTTTTGGAGGC
TGCAGAAGCCTGGAATTCTCAAACAAATGTTTCTTTTTGATAGGTAAGATGAC
TGTGTCTTTCAGATCCGTAATGACATCATCTAAACCTGCTATATCACTCCAAG
TAACATGCATATTAAGAGGGTCTACAAGATGAGCAGCAATACTCATTTCATAT
TTTGAGAGCTTCACATTTTTCACTCCAATTTGCTTCATTAGTTTTTCTGCCTG
TTTCTGAGCTTCTACTTTTTGCTTTCTGGTTGGATCAATTGCATCTACCATCC
ATTTGATAGTAAAGTATGTCACTGCACCAAATATTGTCAAACGGAAAATTAAA
CCAACAACTTCATTCCGACTCAAAGGACGAGAAAAGGCTTCAGCATGTACCAT
CTTGAATGTTAACCCAGGGGCAGAAACAGCCAGAGCAAG

SEQ ID NO. 3:

PROTEIN

LALAVSAPGLTFKMVHAEAFSRPLSRNEVVGLIFRLTIFGAVTYFTIKWMVDA
IDPTRKQKVEAQKQAEKLMKQIGVKNVKLSKYEMSIAAHLVDPLNMHVTWSDI
AGLDDVITDLKDTVILPIKKKHLFENSRLLQPPKGVLLYGPPGWGKTLIAKAT
AKEAGCPFINLQPSTLTDKWYGESQKLAAAVLSLAIKLQPSIIFIDGNRLLFY
ETVQVLTMKATAH

Fig. 2

CF2:

SEQ ID NO. 4:

GATTTAGTAAGTCACCATGTGAGGACAAAACTTGATGAACTGAAAAGGCAAGA
AGTAGGAAGGTTAAGAATGTTAATTAAAGCTAAGTTGGATTCCCTTCAAGATA
TAGGCATGGACCACCAAGCTCTTCTAAAACAATTTGATCACCTAAACCACCTG
AATCCTGACAAGTTTGAATCCACAGATTTAGATATGCTAATCAAAGCGGCAAC
AAGTGATCTGGAACACTATGACAAGACTCGTCATGAAGAATTTAAAAAATATG
AAATGATGAAGGAACATGAAAGGAGAGAATATTTAAAAACATTGAATGAAGAA
AAGAGAAAAGAAGAAGAGTCTAAATTTGAAGAAATGAAGAAAAAGCATGAAAA
TCACCCTAAAGTTAATCACCCAGGAAGCAAAGATCAACTAAAAGAGGTATGGG
AAGAGACTGATGGATTGGATCCTAATGACTTTGACCCCAAGACATTTTTCAAA
TTACATGATGTCAATAGTGATGGATTCCTGGATGAACAAGAATTAGAAGCCCT
ATTTACTAAAGAGTTGGAGAAAGTTTATGACCCTAAAAATGAAAAGGATGATA
TGGTAGAAATGGAAGAAGAAAGGCTTAAAATGAGGGAACATGTAATGAATGAG
GTTGATACTAACAAAGACAGATTGGTGACTCTTGGAGGAGTTTTTGAAAGCCA
CAGAAAAAAAAAGAATTTTTGGAGCCCAGATAGCTGGGA

SEQ ID NO. 5:

REVERSE COMPLEMENT

TCCCAGCTATCTGGGCTCCAAAAATTCTTTTTTTTTTCTGTGGCTTTCAAAAAC
TCCTCCAAGAGTCACCAATCTGTCTTTGTTAGTATCAACCTCATTCATTACAT
GTTCCCTCATTTTAAGCCTTTCTTCTTCCATTTCTACCATATCATCCTTTTCA
TTTTTAGGGTCATAAACTTTCTCCAACTCTTTAGTAAATAGGGCTTCTAATTC
TTGTTCATCCAGGAATCCATCACTATTGACATCATGTAATTTGAAAAATGTCT
TGGGGTCAAAGTCATTAGGATCCAATCCATCAGTCTCTTCCCATACCTCTTTT
AGTTGATCTTTGCTTCCTGGGTGATTAACTTTAGGGTGATTTTCATGCTTTTT
CTTCATTTCTTCAAATTTAGACTCTTCTTCTTTTCTCTTTTCTTCATTCAATG
TTTTTAAATATTCTCTCCTTTCATGTTCCTTCATCATTTCATATTTTTTAAAT
TCTTCATGACGAGTCTTGTCATAGTGTTCCAGATCACTTGTTGCCGCTTTGAT
TAGCATATCTAAATCTGTGGATTCAAACTTGTCAGGATTCAGGTGGTTTAGGT
GATCAAATTGTTTTAGAAGAGCTTGGTGGTCCATGCCTATATCTTGAAGGGAA
TCCAACTTAGCTTTAATTAACATTCTTAACCTTCCTACTTCTTGCCTTTTCAG
TTCATCAAGTTTTGTCCTCACATGGTGACTTACTAAATC

SEQ ID NO. 6:

PROTEIN:

DLVSHHVRTKLDELKRQEVGRLRMLIKAKLDSLQDIGMDHQALLKQFDHLNHL
NPDKFESTDLDMLIKAATSDLEHYDKTRHEEFKKYEMMKEHERREYLKTLNEE
KRKEEESKFEEMKKKHENHPKVNHPGSKDQLKEVWEETDGLDPNDFDPKTFFK
LHDVNSDGFLDEQELEALFTKELEKVYDPKNEKDDMVEMEEERLKMREHVMNE
VDTNKDRLVTLGGVFESHRKKKNFWSPDSW

Fig. 3

CF3:

SEQ ID NO. 7:

TTTGTCAGACACCACGTCCGCACAAAGCTGGATGAGCTCAAGCGACAGGAGGT
GTCACGGCTGCGGATGCTGCTCAAGGCCAAGATGGACGCCGAGCAGGATCCCA
ATGTACAGGTGGATCATCTGAATCTCCTGAAACAGTTTGAACACCTGGACCCT
CAGAACCAGCATACATTCGAGGCCCGCGACCTGGAGCTGCTGATCCAGACGGC
CACCCGGGACCTTGCCCAGTACGACGCAACCCATCATGAAAAGTTCAAGCGCT
ACGAGATGCTTAAGGAACACGAGAGACGGCGTTATCTGGAGTCACTGGGAGAG
GAGCAGAGAAAGGAGGCGGAGAGGAAGCTGGAAGAGCAACAGCGCCGGCACCG
CGAGCACCCTAAAGTCAACGTGCCTGGCAGCCAAGCCCAGTTGAAGGAGGTGT
GGGAGGAGCTGGATGGACTGGACCCCAACAGGTTTAACCCCAAGACCTTCTTC
ATACTGCATGATATCAACAGTGATGGTGTCCTGGATGAGCAGGAGCTGGAGGC
ACTCTTCACCAAGGAGCTGGAGAAAGTGTACGACCCAAAGAATGAGGAGGACG
ACATGCGGGAGATGGAGGAGGAGCGACTGCGCATGCTGAAGCATGTGATGAAG
AATGTGGACACCCAACCAGGACCG

SEQ ID NO. 8:

REVERSE COMPLEMENT:

CGGTCCTGGTTGGGTGTCCACATTCTTCATCACATGCTTCAGCATGCGCAGTC
GCTCCTCCTCCATCTCCCGCATGTCGTCCTCCTCATTCTTTGGGTCGTACACT
TTCTCCAGCTCCTTGGTGAAGAGTGCCTCCAGCTCCTGCTCATCCAGGACACC
ATCACTGTTGATATCATGCAGTATGAAGAAGGTCTTGGGGTTAAACCTGTTGG
GGTCCAGTCCATCCAGCTCCTCCCACACCTCCTTCAACTGGGCTTGGCTGCCA
GGCACGTTGACTTTAGGGTGCTCGCGGTGCCGGCGCTGTTGCTCTTCCAGCTT
CCTCTCCGCCTCCTTTCTCTGCTCCTCTCCCAGTGACTCCAGATAACGCCGTC
TCTCGTGTTCCTTAAGCATCTCGTAGCGCTTGAACTTTTCATGATGGGTTGCG
TCGTACTGGGCAAGGTCCCGGGTGGCCGTCTGGATCAGCAGCTCCAGGTCGCG
GGCCTCGAATGTATGCTGGTTCTGAGGGTCCAGGTGTTCAAACTGTTTCAGGA
GATTCAGATGATCCACCTGTACATTGGGATCCTGCTCGGCGTCCATCTTGGCC
TTGAGCAGCATCCGCAGCCGTGACACCTCCTGTCGCTTGAGCTCATCCAGCTT
TGTGCGGACGTGGTGTCTGACAAA

SEQ ID NO. 9:

PROTEIN:

FVRHHVRTKLDELKRQEVSRLRMLLKAKMDAEQDPNVQVDHLNLLKQFEHLDP
QNQHTFEARDLELLIQTATRDLAQYDATHHEKFKRYEMLKEHERRRYLESLGE
EQRKEAERKLEEQQRRHREHPKVNVPGSQAQLKEVWEELDGLDPNRFNPKTFF
ILHDINSDGVLDEQELEALFTKELEKVYDPKNEEDDMREMEEERLRMLKHVMK
NVDTQPGP

Fig. 4

CF4:

SEQ ID NO. 10:

GGGGACTCGGCCCTGAACGAGCAGGAGAAGGAGTTGCAGCGGCGGCTGAAGCG
TCTCTACCCGGCCGAGGACGAACAAGAGACGCCGCTGCCTAGGTCCTGGAGCC
CGAAGGACAAGTTCAGCTACATCGGCCTCTCTCAGAACAACCTGCGGGTGCAC
TACAAAGGTCATGGCAAAACCCCAAAAGATGCCGCGTCAGTTCGAGCCACGCA
TCCAATACCAGCAGCCTGTGGGATTTATTATTTTGAAGTAAAAATTGTCAGTA
AGGGAAGAGATGGTTACATGGGAATTGGTCTTTCTGCTCAAGGTGTGAACATG
AATAGACTACCAGGTTGGGATAAGCATTCATATGGTTACCATGGGGATGATGG
ACATTCGTTTTGTTCTTCTGGAACTGGACAACCTTATGGACCAACTTTCACTA
CTGGTGATGTCATTGGCTGTTGTGTTAATCTTATCAACAATACCTGCTTTTAC
ACCAAGAATGGACATAGTTTAGGTATTGCTTTACTGACCTACCGCCAAATTTG
TATCCTACTGTGGGGCTTCAAACACCAGGAGAAGTGGTCGATGCCAATTTTTG
GGCAACATCCTTTCCGTGTTTGATATAAAAAACTATATGCCGGGAGTGGAGAA
CCAAAATCCAGGCCCCAGATAGATCCGATTTCCT

SEQ ID NO. 11:

REVERSE COMPLEMENT

AGGAAATCGGATCTATCTGGGGCCTGGATTTTGGTTCTCCACTCCCGGCATAT
AGTTTTTTATATCAAACACGGAAAGGATGTTGCCCAAAAATTGGCATCGACCA
CTTCTCCTGGTGTTTGAAGCCCCACAGTAGGATACAAATTTGGCGGTAGGTCA
GTAAAGCAATACCTAAACTATGTCCATTCTTGGTGTAAAAGCAGGTATTGTTG
ATAAGATTAACACAACAGCCAATGACATCACCAGTAGTGAAAGTTGGTCCATA
AGGTTGTCCAGTTCCAGAAGAACAAAACGAATGTCCATCATCCCCATGGTAAC
CATATGAATGCTTATCCCAACCTGGTAGTCTATTCATGTTCACACCTTGAGCA
GAAAGACCAATTCCCATGTAACCATCTCTTCCCTTACTGACAATTTTTACTTC
AAAATAATAAATCCCACAGGCTGCTGGTATTGGATGCGTGGCTCGAACTGACG
CGGCATCTTTTGGGGTTTTGCCATGACCTTTGTAGTGCACCCGCAGGTTGTTC
TGAGAGAGGCCGATGTAGCTGAACTTGTCCTTCGGGCTCCAGGACCTAGGCAG
CGGCGTCTCTTGTTCGTCCTCGGCCGGGTAGAGACGCTTCAGCCGCCGCTGCA
ACTCCTTCTCCTGCTCGTTCAGGGCCGAGTCCCC

SEQ ID NO. 12:

PROTEIN

GDSALNEQEKELQRRLKRLYPAEDEQETPLPRSWSPKDKFSYIGLSQNNLRVH
YKGHGKTPKDAASVRATHPIPAACGIYYFEVKIVSKGRDGYMGIGLSAQGVNM
NRLPGWDKHSYGYHGDDGHSFCSSGTGQPYGPTFTTGDVIGCCVNLINNTCFY
TKNGHSLGIALLTYRQICILLWGFKHQEKWSMPIFGQHPFRV

Fig. 5

PXR (ORF):

SEQ ID NO. 13:

ATGACATGTGAAGGATGCAAGGGCTTTTTCAGGAGGGCCATGAAACGCAACGC
CCGGCTGAGGTGCCCCTTCCGGAAGGGCGCCTGCGAGATCACCCGGAAGACCC
GGCGACAGTGCCAGGCCTGCCGCCTGCGCAAGTGCCTGGAGAGCGGCATGAAG
AAGGAGATGATCATGTCCGACGAGGCCGTGGAGGAGAGGCGGGCCTTGATCAA
GCGGAAGAAAAGTGAACGGACAGGGACTCAGCCACTGGGAGTGCAGGGGCTGA
CAGAGGAGCAGCGGATGATGATCAGGGAGCTGATGGACGCTCAGATGAAAACC
TTTGACACTACCTTCTCCCATTTCAAGAATTTCCGGCTGCCAGGGGTGCTTAG
CAGTGGCTGCGAGTTGCCAGAGTCTCTGCAGGCCCCATCGAGGGAAGAAGCTG
CCAAGTGGAGCCAGGTCCGGAAAGATCTGTGCTCTTTGAAGGTCTCTCTGCAG
CTGCGGGGGGAGGATGGCAGTGTCTGGAACTACAAACCCCCAGCCGACAGTGG
CGGGAAAGAGATCTTCTCCCTGCTGCCCCACATGGCTGACATGTCAACCTACA
TGTTCAAAGGCATCATCAGCTTTGCCAAAGTCATCTCCTACTTCAGGGACTTG
CCCATCGAGGACCAGATCTCCCTGCTGAAGGGGGCCGCTTTCGAGCTGTGTCA
ACTGAGATTCAACACAGTGTTCAACGCGGAGACTGGAACCTGGGAGTGTGGCC
GGCTGTCCTACTGCTTGGAAGACACTGCAGGTGGCTTCCAGCAACTTCTACTG
GAGCCCATGCTGAAATTCCACTACATGCTGAAGAAGCTGCAGCTGCATGAGGA
GGAGTATGTGCTGATGCAGGCCATCTCCCTCTTCTCCCCAGACCGCCCAGGTG
TGCTGCAGCACCGCGTGGTGGACCAGCTGCAGGAGCAATTCGCCATTACTCTG
AAGTCCTACATTGAATGCAATCGGCCCCAGCCTGCTCATAGGTTCTTGTTCCT
GAAGATCATGGCTATGCTCACCGAGCTCCGCAGCATCAATGCTCAGCACACCC
AGCGGCTGCTGCGCATCCAGGACATACACCCCTTTGCTACGCCCCTCATGCAG
GAGTTGTTCGGCATCACAGGTAGCTGA

PXR Reverse complement:

SEQ ID NO. 14:

AGTCGATGGACACTACGGCTTGTTGAGGACGTACTCCCCGCATCGTTTCCCCA
CATACAGGACCTACGCGTCGTCGGCGACCCACACGACTCGTAACTACGACGCC
TCGAGCCACTCGTATCGGTACTAGAAGTCCTTGTTCTTGGATACTCGTCCGAC
CCCGGCTAACGTAAGTTACATCCTGAAGTCTCATTACCGCTTAACGAGGACGT
CGACCAGGTGGTGCGCCACGACGTCGTGTGGACCCGCCAGACCCCTCTTCTCC
CTCTACCGGACGTAGTCGTGTATGAGGAGGAGTACGTCGACGTCGAAGAAGTC
GTACATCACCTTAAAGTCGTACCCGAGGTCATCTTCAACGACCTTCGGTGGAC
GTCACAGAAGGTTCGTCATCCTGTCGGCCGGTGTGAGGGTCCAAGGTCAGAGG
CGCAACTTGTGACACAACTTAGAGTCAACTGTGTCGAGCTTTCGCCGGGGGAA
GTCGTCCCTCTAGACCAGGAGCTACCCGTTCAGGGACTTCATCCTCTACTGAA
ACCGTTTCGACTACTACGGAAACTTGTACATCCAACTGTACAGTCGGTACACC
CCGTCGTCCCTCTTCTAGAGAAAGGGCGGTGACAGCCGACCCCCAAACATCAA
GGTCTGTGACGGTAGGAGGGGGGCGTCGACGTCTCTCTGGAAGTTTCTCGTGT
CTAGAAAGGCCTGGACCGAGGTGAACCGTCGAAGAAGGGAGCTACCCCGGACG
TCTCTGAGACCGTTGAGCGTCGGTGACGATTCGTGGGGACCGTCGGCCTTTAA
GAACTTTACCCTCTTCCATCACAGTTTCCAAAAGTAGACTCGCAGGTAGTCGA
GGGACTAGTAGTAGGCGACGAGGAGACAGTCGGGGACGTGAGGGTCACCGACT
CAGGGACAGGCAAGTGAAAAGAAGGCGAACTAGTTCCGGGCGGAGAGGAGGTG
CCGGAGCAGCCTGTACTAGTAGAGGAAGAAGTACGGCGAGAGGTCCGTGAACG

CGTCCGCCGTCCGGACCGTGACAGCGGCCCAGAAGGCCCACTAGAGCGTCCGC
GGGAAGGCCTTCCCCGTGGAGTCGGCCCGCAACGCAAAGTACCGGGAGGACTT
TTTCGGGAACGTAGGAAGTGTACAGTA

PXR-Protein

SEQ ID NO. 15:

MEVRPKESWNHADFVHCEDTESVPGKPSVNADEEVGGPQICRVCGDKATGYHF
NVMTCEGCKGFFRRAMKRNARLRCPFRKGACEITRKTRRQCQACRLRKCLESG
MKKEMIMSDEAVEERRALIKRKKSERTGTQPLGVQGLTEEQRMMIRELMDAQM
KTFDTTFSHFKNFRLPGVLSSGCELPESLQAPSREEAAKWSQVRKDLCSLKVS
LQLRGEDGSVWNYKPPADSGGKEIFSLLPHMADMSTYMFKGIISFAKVISYFR
DLPIEDQISLLKGAAFELCQLRFNTVFNAETGTWECGRLSYCLEDTAGGFQQL
LLEPMLKFHYMLKKLQLHEEEYVLMQAISLFSPDRPGVLQHRVVDQLQEQFAI
TLKSYIECNRPQPAHRFLFLKIMAMLTELRSINAQHTQRLLRIQDIHPFATPL
MQELFGITGS

Fig. 6

PXR (Ligand Binding Domain LBD)

SEQ ID NO. 16:

GGCATGAAGAAGGAGATGATCATGTCCGACGAGGCCGTGGAGGAGAGGCGGGC
CTTGATCAAGCGGAAGAAAAGTGAACGGACAGGGACTCAGCCACTGGGAGTGC
AGGGGCTGACAGAGGAGCAGCGGATGATGATCAGGGAGCTGATGGACGCTCAG
ATGAAAACCTTTGACACTACCTTCTCCCATTTCAAGAATTTCCGGCTGCCAGG
GGTGCTTAGCAGTGGCTGCGAGTTGCCAGAGTCTCTGCAGGCCCCATCGAGGG
AAGAAGCTGCCAAGTGGAGCCAGGTCCGGAAAGATCTGTGCTCTTTGAAGGTC
TCTCTGCAGCTGCGGGGGGAGGATGGCAGTGTCTGGAACTACAAACCCCCAGC
CGACAGTGGCGGGAAAGAGATCTTCTCCCTGCTGCCCCACATGGCTGACATGT
CAACCTACATGTTCAAAGGCATCATCAGCTTTGCCAAAGTCATCTCCTACTTC
AGGGACTTGCCCATCGAGGACCAGATCTCCCTGCTGAAGGGGGCCGCTTTCGA
GCTGTGTCAACTGAGATTCAACACAGTGTTCAACGCGGAGACTGGAACCTGGG
AGTGTGGCCGGCTGTCCTACTGCTTGGAAGACACTGCAGGTGGCTTCCAGCAA
CTTCTACTGGAGCCCATGCTGAAATTCCACTACATGCTGAAGAAGCTGCAGCT
GCATGAGGAGGAGTATGTGCTGATGCAGGCCATCTCCCTCTTCTCCCCAGACC
GCCCAGGTGTGCTGCAGCACCGCGTGGTGGACCAGCTGCAGGAGCAATTCGCC
ATTACTCTGAAGTCCTACATTGAATGCAATCGGCCCCAGCCTGCTCATAGGTT
CTTGTTCCTGAAGATCATGGCTATGCTCACCGAGCTCCGCAGCATCAATGCTC
AGCACACCCAGCGGCTGCTGCGCATCCAGGACATACACCCCTTTGCTACGCCC
CTCATGCAGGAGTTGTTCGGCATCACAGGTAGCTGA

PXR-LBD reverse complement

SEQ ID NO. 17:

TCAGCTACCTGTGATGCCGAACAACTCCTGCATGAGGGGCGTAGCAAAGGGGT
GTATGTCCTGGATGCGCAGCAGCCGCTGGGTGTGCTGAGCATTGATGCTGCGG
AGCTCGGTGAGCATAGCCATGATCTTCAGGAACAAGAACCTATGAGCAGGCTG
GGGCCGATTGCATTCAATGTAGGACTTCAGAGTAATGGCGAATTGCTCCTGCA
GCTGGTCCACCACGCGGTGCTGCAGCACACCTGGGCGGTCTGGGGAGAAGAGG
GAGATGGCCTGCATCAGCACATACTCCTCCTCATGCAGCTGCAGCTTCTTCAG
CATGTAGTGGAATTTCAGCATGGGCTCCAGTAGAAGTTGCTGGAAGCCACCTG
CAGTGTCTTCCAAGCAGTAGGACAGCCGGCCACACTCCCAGGTTCCAGTCTCC
GCGTTGAACACTGTGTTGAATCTCAGTTGACACAGCTCGAAAGCGGCCCCCTT
CAGCAGGGAGATCTGGTCCTCGATGGGCAAGTCCCTGAAGTAGGAGATGACTT
TGGCAAAGCTGATGATGCCTTTGAACATGTAGGTTGACATGTCAGCCATGTGG
GGCAGCAGGGAGAAGATCTCTTTCCCGCCACTGTCGGCTGGGGGTTTGTAGTT
CCAGACACTGCCATCCTCCCCCCGCAGCTGCAGAGAGACCTTCAAAGAGCACA
GATCTTTCCGGACCTGGCTCCACTTGGCAGCTTCTTCCCTCGATGGGGCCTGC
AGAGACTCTGGCAACTCGCAGCCACTGCTAAGCACCCCTGGCAGCCGGAAATT
CTTGAAATGGGAGAAGGTAGTGTCAAAGGTTTTCATCTGAGCGTCCATCAGCT
CCCTGATCATCATCCGCTGCTCCTCTGTCAGCCCCTGCACTCCCAGTGGCTGA
GTCCCTGTCCGTTCACTTTTCTTCCGCTTGATCAAGGCCCGCCTCTCCTCCAC
GGCCTCGTCGGACATGATCATCTCCTTCTTCATGCC

PXR-LBD Protein:

SEQ ID NO. 18:

GMKKEMIMSDEAVEERRALIKRKKSERTGTQPLGVQGLTEEQRMMIRELMDAQ
MKTFDTTFSHFKNFRLPGVLSSGCELPESLQAPSREEAAKWSQVRKDLCSLKV
SLQLRGEDGSVWNYKPPADSGGKEIFSLLPHMADMSTYMFKGIISFAKVISYF
RDLPIEDQISLLKGAAFELCQLRFNTVFNAETGTWECGRLSYCLEDTAGGFQQ
LLLEPMLKFHYMLKKLQLHEEEYVLMQAISLFSPDRPGVLQHRVVDQLQEQFA
ITLKSYIECNRPQPAHRFLFLKIMAMLTELRSINAQHTQRLLRIQDIHPFATP
LMQELFGITGS

Fig. 7

RXRalpha (ORF):

SEQ ID NO. 19:

ATGGACACCAAACATTTCCTGCCGCTCGATTTCTCCACCCAGGTGAACTCCTC
CCTCACCTCCCCGACGGGGCGAGGCTCCATGGCTGCCCCCTCGCTGCACCCGT
CCCTGGGGCCTGGCATCGGCTCCCCGGGACAGCTGCATTCTCCCATCAGCACC
CTGAGCTCCCCCATCAACGGCATGGGCCCGCCTTTCTCGGTCATCAGCTCCCC
CATGGGCCCCCACTCCATGTCGGTGCCCACCACACCCACCCTGGGCTTCAGCA
CTGGCAGCCCCCAGCTCAGCTCACCTATGAACCCCGTCAGCAGCAGCGAGGAC
ATCAAGCCCCCCCTGGGCCTCAATGGCGTCCTCAAGGTCCCCGCCCACCCCTC
AGGAAACATGGCTTCCTTCACCAAGCACATCTGCGCCATCTGCGGGGACCGCT
CCTCAGGCAAGCACTATGGAGTGTACAGCTGCGAGGGGTGCAAGGGCTTCTTC
AAGCGGACGGTGCGCAAGGACCTGACCTACACCTGCCGCGACAACAAGGACTG
CCTGATTGACAAGCGGCAGCGGAACCGGTGCCAGTACTGCCGCTACCAGAAGT
GCCTGGCCATGGGCATGAAGCGGGAAGCCGTGCAGGAGGAGCGGCAGCGTGGC
AAGGACCGGAACGAGAATGAGGTGGAGTCGACCAGCAGCGCCAACGAGGACAT
GCCGGTGGAGAGGATCCTGGAGGCTGAGCTGGCCGTGGAGCCCAAGACCGAGA
CCTACGTGGAGGCAAACATGGGGCTGAACCCCAGCTCGCCGAACGACCCTGTC
ACCAACATTTGCCAAGCAGCCGACAAACAGCTTTTCACCCTGGTGGAGTGGGC
CAAGCGGATCCCACACTTCTCAGAGCTGCCCCTGGACGACCAGGTCATCCTGC
TGCGGGCAGGCTGGAATGAGCTGCTCATCGCCTCCTTCTCCCACCGCTCCATC
GCCGTGAAGGACGGGATCCTCCTGGCCACCGGGCTGCACGTCCACCGGAACAG
CGCCCACAGCGCAGGGGTGGGCGCCATCTTTGACAGGGTGCTGACGGAGCTTG
TGTCCAAGATGCGGGACATGCAGATGGACAAGACGGAGCTGGGCTGCCTGCGC
GCCATCGTCCTCTTTAACCCTGACTCCAAGGGGCTCTCGAACCCGGCCGAGGT
GGAGGCGCTGAGGGAGAAGGTCTATGCGTCCTTGGAGGCCTACTGCAAGCACA
AGTACCCAGAGCAGCCGGGAAGGTTCGCTAAGCTCTTGCTCCGCCTGCCGGCT
CTGCGCTCCATCGGGCTCAAATGCCTGGAACATCTCTTCTTCTTCAAGCTCAT
CGGGGACACACCCATTGACACCTTCCTTATGGAGATGCTGGAGGCGCCGCACC
AAATGACTTAG

RXRalpha reverse complement:

SEQ ID NO. 20:

CTAAGTCATTTGGTGCGGCGCCTCCAGCATCTCCATAAGGAAGGTGTCAATGG
GTGTGTCCCCGATGAGCTTGAAGAAGAAGAGATGTTCCAGGCATTTGAGCCCG
ATGGAGCGCAGAGCCGGCAGGCGGAGCAAGAGCTTAGCGAACCTTCCCGGCTG
CTCTGGGTACTTGTGCTTGCAGTAGGCCTCCAAGGACGCATAGACCTTCTCCC
TCAGCGCCTCCACCTCGGCCGGGTTCGAGAGCCCCTTGGAGTCAGGGTTAAAG
AGGACGATGGCGCGCAGGCAGCCCAGCTCCGTCTTGTCCATCTGCATGTCCCG
CATCTTGGACACAAGCTCCGTCAGCACCCTGTCAAAGATGGCGCCCACCCCTG
CGCTGTGGGCGCTGTTCCGGTGGACGTGCAGCCCGGTGGCCAGGAGGATCCCG
TCCTTCACGGCGATGGAGCGGTGGGAGAAGGAGGCGATGAGCAGCTCATTCCA
GCCTGCCCGCAGCAGGATGACCTGGTCGTCCAGGGGCAGCTCTGAGAAGTGTG
GGATCCGCTTGGCCCACTCCACCAGGGTGAAAAGCTGTTTGTCGGCTGCTTGG
CAAATGTTGGTGACAGGGTCGTTCGGCGAGCTGGGGTTCAGCCCCATGTTTGC
CTCCACGTAGGTCTCGGTCTTGGGCTCCACGGCCAGCTCAGCCTCCAGGATCC
TCTCCACCGGCATGTCCTCGTTGGCGCTGCTGGTCGACTCCACCTCATTCTCG

TTCCGGTCCTTGCCACGCTGCCGCTCCTCCTGCACGGCTTCCCGCTTCATGCC
CATGGCCAGGCACTTCTGGTAGCGGCAGTACTGGCACCGGTTCCGCTGCCGCT
TGTCAATCAGGCAGTCCTTGTTGTCGCGGCAGGTGTAGGTCAGGTCCTTGCGC
ACCGTCCGCTTGAAGAAGCCCTTGCACCCCTCGCAGCTGTACACTCCATAGTG
CTTGCCTGAGGAGCGGTCCCCGCAGATGGCGCAGATGTGCTTGGTGAAGGAAG
CCATGTTTCCTGAGGGGTGGGCGGGGACCTTGAGGACGCCATTGAGGCCCAGG
GGGGGCTTGATGTCCTCGCTGCTGCTGACGGGGTTCATAGGTGAGCTGAGCTG
GGGGCTGCCAGTGCTGAAGCCCAGGGTGGGTGTGGTGGGCACCGACATGGAGT
GGGGGCCCATGGGGGAGCTGATGACCGAGAAAGGCGGGCCCATGCCGTTGATG
GGGGAGCTCAGGGTGCTGATGGGAGAATGCAGCTGTCCCGGGGAGCCGATGCC
AGGCCCCAGGGACGGGTGCAGCGAGGGGGCAGCCATGGAGCCTCGCCCCGTCG
GGGAGGTGAGGGAGGAGTTCACCTGGGTGGAGAAATCGAGCGGCAGGAAATGT
TTGGTGTCCAT

RXRalpha-Protein.

SEQ ID NO. 21:

MDTKHFLPLDFSTQVNSSLTSPTGRGSMAAPSLHPSLGPGIGSPGQLHSPIST
LSSPINGMGPPFSVISSPMGPHSMSVPTTPTLGFSTGSPQLSSPMNPVSSSED
IKPPLGLNGVLKVPAHPSGNMASFTKHICAICGDRSSGKHYGVYSCEGCKGFF
KRTVRKDLTYTCRDNKDCLIDKRQRNRCQYCRYQKCLAMGMKREAVQEERQRG
KDRNENEVESTSSANEDMPVERILEAELAVEPKTETYVEANMGLNPSSPNDPV
TNICQAADKQLFTLVEWAKRIPHFSELPLDDQVILLRAGWNELLIASFSHRSI
AVKDGILLATGLHVHRNSAHSAGVGAIFDRVLTELVSKMRDMQMDKTELGCLR
AIVLFNPDSKGLSNPAEVEALREKVYASLEAYCKHKYPEQPGRFAKLLLRLPA
LRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPHQMT

Fig. 8

RXRb ta (ORF):

SEQ ID NO. 22:

ATGTCTTGGGCCGCTCGCCCGCCCTTCCTCCCTCAGCGGCATGCCGCAGGGCA
GTGTGGGCCGGTGGGGGTGCGAAAAGAAATGCATTGTGGGGTCGCGTCCCGGT
GGCGGCGGCGACGGCCCTGGCTGGATCCCGCAGCGGCGGCGGCGGCGGCGGTG
GCAGGCGGAGAACAACAAACCCCGGAGCCGGAGCCAGGGGAGGCTGGACGGGA
CGGGATGGGCGACAGCGGGCGGGACTCCCGAAGCCCAGACAGCTCCTCCCCAA
ATCCCCTTCCCCAGGGAGTCCCTCCCCCTTCTCCTCCTGGGCCACCCCTACCC
CCTTCAACAGCTCCTACCCTTGGAGGCTCTGGGGCCCCACCCCCACCCCCGAT
GCCACCACCCCCACTGGGCTCTCCCTTTCCAGTCATCAGTTCTTCCATGGGGT
CCCCTGGTCTGCCCCCTCCAGCTCCCCCAGGATTCTCCGGGCCTGTCAGCAGC
CCCCAGATTAACTCAACAGTGTCACTCCCTGGGGGTGGGTCTGGCCCCCCTGA
AGATGTGAAGCCACCAGTCTTAGGGGTCCGGGGCCTGCACTGTCCACCCCCTC
CAGGTGGCCCTGGGGCTGGCAAACGGCTATGTGCAATCTGCGGGGACAGAAGC
TCAGGCAAACACTACGGGGTTTACAGCTGTGAGGGTTGCAAGGGCTTCTTCAA
ACGCACCATCCGCAAAGACCTTACATACTCTTGCCGGGACAACAAAGACTGCA
CAGTGGACAAGCGCCAGCGGAACCGCTGTCAGTACTGCCGCTATCAGAAGTGC
CTGGCCACTGGCATGAAGAGGGAGGCGGTACAGGAGGAGCGTCAGCGGGGAAA
GGACAAGGATGGGGATGGGGAGGGGGCTGGGGGAGCCCCCGAGGAGATGCCTG
TGGACAGGATCCTGGAGGCAGAGCTTGCTGTGGAACAGAAGAGTGACCAGGGC
GTTGAGGGTCCTGGGGGGAACCGGGGGTAGCGGCAGCAGCCCAAATGACCCTGT
GACTAACATCTGTCAGGCAGCTGACAAACAGCTATTCACGCTTGTTGAGTGGG
CGAAGAGGATCCCACACTTTTCCTCCTTGCCTCTGGATGATCAGGTCATATTG
CTGCGGGCAGGCTGGAATGAACTCCTCATTGCCTCCTTTTCACACCGATCCAT
TGATGTTCGAGATGGCATCCTCCTTGCCACAGGTCTTCACGTGCACCGCAACT
CAGCCCATTCAGCAGGAGTAGGAGCCATCTTTGATCGGGTGCTGACAGAGCTA
GTGTCCAAAATGCGTGACATGAGGATGGACAAGACAGAGCTTGGCTGCCTGAG
GGCAATCATTCTGTTTAATCCAGATGCCAAGGGCCTCTCCAACCCTAGTGAGG
TGGAGGTCCTGCGGGAGAAAGTGTATGCATCACTGGAGACCTACTGCAAACAG
AAGTACCCTGAGCAGCAGGGACGGTTTGCCAAGCTGCTGCTACGTCTTCCTGC
CCTCCGGTCCATTGGCCTTAAGTGTCTAGAGCATCTGTTTTTCTTCAAGCTCA
TTGGTGACACCCCCATCGACACCTTCCTCATGGAGATGCTTGAGGCTCCCCAT
CAACTGGCCTGA

RXRbeta reverse complement:

SEQ ID NO. 23:

TCAGGCCAGTTGATGGGGAGCCTCAAGCATCTCCATGAGGAAGGTGTCGATGG
GGGTGTCACCAATGAGCTTGAAGAAAAACAGATGCTCTAGACACTTAAGGCCA
ATGGACCGGAGGGCAGGAAGACGTAGCAGCAGCTTGGCAAACCGTCCCTGCTG
CTCAGGGTACTTCTGTTTGCAGTAGGTCTCCAGTGATGCATACACTTTCTCCC
GCAGGACCTCCACCTCACTAGGGTTGGAGAGGCCCTTGGCATCTGGATTAAAC
AGAATGATTGCCCTCAGGCAGCCAAGCTCTGTCTTGTCCATCCTCATGTCACG
CATTTTGGACACTAGCTCTGTCAGCACCCGATCAAAGATGGCTCCTACTCCTG
CTGAATGGGCTGAGTTGCGGTGCACGTGAAGACCTGTGGCAAGGAGGATGCCA
TCTCGAACATCAATGGATCGGTGTGAAAAGGAGGCAATGAGGAGTTCATTCCA
GCCTGCCCGCAGCAATATGACCTGATCATCCAGAGGCAAGGAGGAAAAGTGTG

GGATCCTCTTCGCCCACTCAACAAGCGTGAATAGCTGTTTGTCAGCTGCCTGA
CAGATGTTAGTCACAGGGTCATTTGGGCTGCTGCCGCTACCCCCGGTTCCCCC
AGGACCCTCAACGCCCTGGTCACTCTTCTGTTCCACAGCAAGCTCTGCCTCCA
GGATCCTGTCCACAGGCATCTCCTCGGGGGCTCCCCCAGCCCCCTCCCCATCC
CCATCCTTGTCCTTTCCCCGCTGACGCTCCTCCTGTACCGCCTCCCTCTTCAT
GCCAGTGGCCAGGCACTTCTGATAGCGGCAGTACTGACAGCGGTTCCGCTGGC
GCTTGTCCACTGTGCAGTCTTTGTTGTCCCGGCAAGAGTATGTAAGGTCTTTG
CGGATGGTGCGTTTGAAGAAGCCCTTGCAACCCTCACAGCTGTAAACCCCGTA
GTGTTTGCCTGAGCTTCTGTCCCCGCAGATTGCACATAGCCGTTTGCCAGCCC
CAGGGCCACCTGGAGGGGGTGGACAGTGCAGGCCCCGGACCCCTAAGACTGGT
GGCTTCACATCTTCAGGGGGGCCAGACCCACCCCCAGGGAGTGACACTGTTGA
GTTAATCTGGGGGCTGCTGACAGGCCCGGAGAATCCTGGGGGAGCTGGAGGGG
GCAGACCAGGGGACCCCATGGAAGAACTGATGACTGGAAAGGGAGAGCCCAGT
GGGGGTGGTGGCATCGGGGGTGGGGGTGGGGCCCCAGAGCCTCCAAGGGTAGG
AGCTGTTGAAGGGGGTAGGGGTGGCCCAGGAGGAGAAGGGGGAGGGACTCCCT
GGGGAAGGGGATTTGGGGAGGAGCTGTCTGGGCTTCGGGAGTCCCGCCCGCTG
TCGCCCATCCCGTCCCGTCCAGCCTCCCCTGGCTCCGGCTCCGGGGTTTGTTG
TTCTCCGCCTGCCACCGCCGCCGCCGCCGCCGCTGCGGGATCCAGCCAGGGCC
GTCGCCGCCGCCACCGGGACGCGACCCCACAATGCATTTCTTTTCGCACCCCC
ACCGGCCCACACTGCCCTGCGGCATGCCGCTGAGGGAGGAAGGGCGGGCGAGC
GGCCCAAGACAT

RXRbeta-Protein:

SEQ ID NO. 24:

MSWAARPPFLPQRHAAGQCGPVGVRKEMHCGVASRWRRRRPWLDPAAAAAAAV
AGGEQQTPEPEPGEAGRDGMGDSGRDSRSPDSSSPNPLPQGVPPPSPPGPPLP
PSTAPTLGGSGAPPPPPMPPPPLGSPFPVISSSMGSPGLPPPAPPGFSGPVSS
PQINSTVSLPGGGSGPPEDVKPPVLGVRGLHCPPPPGGPGAGKRLCAICGDRS
SGKHYGVYSCEGCKGFFKRTIRKDLTYSCRDNKDCTVDKRQRNRCQYCRYQKC
LATGMKREAVQEERQRGKDKDGDGEGAGGAPEEMPVDRILEAELAVEQKSDQG
VEGPGGTGGSGSSPNDPVTNICQAADKQLFTLVEWAKRIPHFSSLPLDDQVIL
LRAGWNELLIASFSHRSIDVRDGILLATGLHVHRNSAHSAGVGAIFDRVLTEL
VSKMRDMRMDKTELGCLRAIILFNPDAKGLSNPSEVEVLREKVYASLETYCKQ
KYPEQQGRFAKLLLRLPALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLEAPH
QLA

Fig. 9

RXRgamma (ORF):

SEQ ID NO. 25:

ATGTATGGAAATTATTCTCACTTCATGAAGTTTCCCGCAGGCTATGGAGGCTC
CCCTGGCCACACTGGCTCTACATCCATGAGCCCATCAGCAGCCTTGTCCACAG
GGAAGCCAATGGACAGCCACCCCAGCTACACAGATACCCCAGTGAGTGCCCCA
CGGACTCTGAGTGCAGTGGGGACCCCCCTCAATGCCCTGGGCTCTCCATATCG
AGTCATCACCTCTGCCATGGGCCCACCCTCAGGAGCACTTGCAGCGCCTCCAG
GAATCAACTTGGTTGCCCCACCCAGCTCTCAGCTAAATGTGGTCAACAGTGTC
AGCAGTTCAGAGGACATCAAGCCCTTACCAGGGCTTCCCGGGATTGGAAACAT
GAACTACCCATCCACCAGCCCCGGATCTCTGGTTAAACACATCTGTGCTATCT
GTGGAGACAGATCCTCAGGAAAGCACTACGGGGTATACAGTTGTGAAGGCTGC
AAAGGGTTCTTCAAGAGGACGATAAGGAAGGACCTCATCTACACGTGTCGGGA
TAATAAAGACTGCCTCATTGACAAGCGTCAGCGCAACCGCTGCCAGTACTGTC
GCTATCAGAAGTGCCTTGTCATGGGCATGAAGAGGGAAGCTGTGCAAGAAGAA
AGACAGAGGAGCCGAGAGCGAGCTGAGAGTGAGGCAGAATGTGCTACCAGTGG
TCATGAAGACATGCCTGTGGAGAGGATTCTAGAAGCTGAACTTGCTGTTGAAC
CAAAGACAGAATCCTATGGTGACATGAATATGGAGAACTCGACAAATGACCCT
GTTACCAACATATGTCATGCTGCTGACAAGCAGCTTTTCACCCTCGTTGAATG
GGCCAAGCGTATTCCCCACTTCTCTGACCTCACCTTGGAGGACCAGGTCATTT
TGCTTCGGGCAGGGTGGAATGAATTGCTGATTGCCTCTTTCTCCCACCGCTCA
GTTTCCGTGCAGGATGGCATCCTTCTGGCCACGGGTTTACATGTCCACCGGAG
CAGTGCCCACAGTGCTGGGGTCGGCTCCATCTTTGACAGAGTTCTAACTGAGC
TGGTTTCCAAAATGAAAGACATGCAGATGGACAAGTCGGAACTGGGATGCCTG
CGAGCCATTGTACTCTTTAACCCAGATGCCAAGGGCCTGTCCAACCCCTCTGA
GGTGGAGACTCTGCGAGAGAAGGTTTATGCCACCCTTGAGGCCTACACCAAGC
AGAAGTATCCGGAACAGCCAGGCAGGTTTGCCAAGCTGCTGCTGCGCCTCCCA
GCTCTGCGTTCCATTGGCTTGAAATGCCTGGAGCACCTCTTCTTCTTCAAGCT
CATCGGGGACACCCCCATTGACACCTTCCTCATGGAGATGTTGGAGACCCCGC
TGCAGATCACCTGA

RXRgamma reverse complement:

SEQ ID NO. 26:

TCAGGTGATCTGCAGCGGGGTCTCCAACATCTCCATGAGGAAGGTGTCAATGG
GGGTGTCCCCGATGAGCTTGAAGAAGAAGAGGTGCTCCAGGCATTTCAAGCCA
ATGGAACGCAGAGCTGGGAGGCGCAGCAGCAGCTTGGCAAACCTGCCTGGCTG
TTCCGGATACTTCTGCTTGGTGTAGGCCTCAAGGGTGGCATAAACCTTCTCTC
GCAGAGTCTCCACCTCAGAGGGGTTGGACAGGCCCTTGGCATCTGGGTTAAAG
AGTACAATGGCTCGCAGGCATCCCAGTTCCGACTTGTCCATCTGCATGTCTTT
CATTTTGGAAACCAGCTCAGTTAGAACTCTGTCAAAGATGGAGCCGACCCCAG
CACTGTGGGCACTGCTCCGGTGGACATGTAAACCCGTGGCCAGAAGGATGCCA
TCCTGCACGGAAACTGAGCGGTGGGAGAAAGAGGCAATCAGCAATTCATTCCA
CCCTGCCCGAAGCAAAATGACCTGGTCCTCCAAGGTGAGGTCAGAGAAGTGGG
GAATACGCTTGGCCCATTCAACGAGGGTGAAAAGCTGCTTGTCAGCAGCATGA
CATATGTTGGTAACAGGGTCATTTGTCGAGTTCTCCATATTCATGTCACCATA
GGATTCTGTCTTTGGTTCAACAGCAAGTTCAGCTTCTAGAATCCTCTCCACAG
GCATGTCTTCATGACCACTGGTAGCACATTCTGCCTCACTCTCAGCTCGCTCT

CGGCTCCTCTGTCTTTCTTCTTGCACAGCTTCCCTCTTCATGCCCATGACAAG
GCACTTCTGATAGCGACAGTACTGGCAGCGGTTGCGCTGACGCTTGTCAATGA
GGCAGTCTTTATTATCCCGACACGTGTAGATGAGGTCCTTCCTTATCGTCCTC
TTGAAGAACCCTTTGCAGCCTTCACAACTGTATACCCCGTAGTGCTTTCCTGA
GGATCTGTCTCCACAGATAGCACAGATGTGTTTAACCAGAGATCCGGGGCTGG
TGGATGGGTAGTTCATGTTTCCAATCCCGGGAAGCCCTGGTAAGGGCTTGATG
TCCTCTGAACTGCTGACACTGTTGACCACATTTAGCTGAGAGCTGGGTGGGGC
AACCAAGTTGATTCCTGGAGGCGCTGCAAGTGCTCCTGAGGGTGGGCCCATGG
CAGAGGTGATGACTCGATATGGAGAGCCCAGGGCATTGAGGGGGGTCCCCACT
GCACTCAGAGTCCGTGGGGCACTCACTGGGGTATCTGTGTAGCTGGGGTGGCT
GTCCATTGGCTTCCCTGTGGACAAGGCTGCTGATGGGCTCATGGATGTAGAGC
CAGTGTGGCCAGGGGAGCCTCCATAGCCTGCGGGAAACTTCATGAAGTGAGAA
TAATTTCCATACAT

RXRgamma-Protein:

SEQ ID NO. 27:

MYGNYSHFMKFPAGYGGSPGHTGSTSMSPSAALSTGKPMDSHPSYTDTPVSAP
RTLSAVGTPLNALGSPYRVITSAMGPPSGALAAPPGINLVAPPSSQLNVVNSV
SSSEDIKPLPGLPGIGNMNYPSTSPGSLVKHICAICGDRSSGKHYGVYSCEGC
KGFFKRTIRKDLIYTCRDNKDCLIDKRQRNRCQYCRYQKCLVMGMKREAVQEE
RQRSRERAESEAECATSGHEDMPVERILEAELAVEPKTESYGDMNMENSTNDP
VTNICHAADKQLFTLVEWAKRIPHFSDLTLEDQVILLRAGWNELLIASFSHRS
VSVQDGILLATGLHVHRSSAHSAGVGSIFDRVLTELVSKMKDMQMDKSELGCL
RAIVLFNPDAKGLSNPSEVETLREKVYATLEAYTKQKYPEQPGRFAKLLLRLP
ALRSIGLKCLEHLFFFKLIGDTPIDTFLMEMLETPLQIT

Fig. 10

CF 44:

CDNA Sequence

SEQ ID NO. 28:

GACTCCCAAGATGGCGGACCTACTGGGCTCCATCCTGAGCTCCATGGAGAAGC
CACCCAGCCTCGGTGACCAGGAGACTCGGCGCAAGGCCCGAGAACAGGCCGCC
CGCCTGAAGAAACTACAAGAGCAAGAGAAACAACAGAAAGTGGAGTTTCGTAA
AAGGATGGAGAAGGAGGTGTCAGATTTCATTCAAGACAGTGGGCAGATCAAGA
AAAAGTTTCAGCCAATGAACAAGATCGAGAGGAGCATACTACATGATGTGGTG
GAAGTGGCTGGCCTGACATCCTTCTCCTTTGGGGAAGATGATGACTGTCGCTA
TGTCATGATCTTCAAAAAGGAGTTTGCACCCTCAGATGAAGAGCTAGACTCTT
ACCGTCGTGGAGAGGAATGGGACCCCCAGAAGGCTGAGGAGAAGCGGAAGCTG
AAGGAGCTGGCCCAGAGGCAAGAGGAGGAGGCAGCCCAGCAGGGGCCTGTGGT
GGTGAGCCCTGCCAGCGACTACAAGGACAAGTACAGCCACCTCATCGGCAAGG
GAGCAGCCAAAGACGCAGCCCACATGCTACAGGCCAATAAGACCTACGGCTGT
GTGCCCGTGGCCAATAAGAGGGACACACGCTCCATTGAAGAGGCTATGAATGA
GATCAGAGCCAAGAAGCGTCTGCGGCAGAGTGGGGAAGAGTTGCCGCCAACCT
CTAGGCGCCCCGCCCAGCTCCCTTTGACCCCTGGGGCAGGGCAGGGGGCAGGG
AGAGACAAGGCTGCTGCTATTAGAGCCCATCCTGGAGCCCCACCTCTGAACCA
CCTCCTACCAGCTGTCCCTCAGGCTGGGGGAAAACAGGTGTTTGATTTGTCAC
CGTTGGAGCTTGGATATGTGCGTGGCATGTGTGTGTGTGTGTGAGAGTGTGAA
TGCACAGGTGGGTATTTAATCTGTATTATTCCCCGTTCTTGGAATTTTCTTCC
CCATGGGGCTGGGGTACTTTACATTCAATAAATACTGTTTAACCCAAAAAAAA
AAAAAAAAAAAAAAAAAAA

SEQ ID NO. 29:

Reverse complement:

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGTTAAACAGTATTTATTGAATGT
AAAGTACCCCAGCCCCATGGGGAAGAAAATTCCAAGAACGGGGAATAATACAG
ATTAAATACCCACCTGTGCATTCACACTCTCACACACACACACACATGCCACG
CACATATCCAAGCTCCAACGGTGACAAATCAAACACCTGTTTTCCCCCAGCCT
GAGGGACAGCTGGTAGGAGGTGGTTCAGAGGTGGGGCTCCAGGATGGGCTCTA
ATAGCAGCAGCCTTGTCTCTCCCTGCCCCCTGCCCTGCCCCAGGGGTCAAAGG
GAGCTGGGCGGGGCGCCTAGAGGTTGGCGGCAACTCTTCCCCACTCTGCCGCA
GACGCTTCTTGGCTCTGATCTCATTCATAGCCTCTTCAATGGAGCGTGTGTCC
CTCTTATTGGCCACGGGCACACAGCCGTAGGTCTTATTGGCCTGTAGCATGTG
GGCTGCGTCTTTGGCTGCTCCCTTGCCGATGAGGTGGCTGTACTTGTCCTTGT
AGTCGCTGGCAGGGCTCACCACCACAGGCCCCTGCTGGGCTGCCTCCTCCTCT
TGCCTCTGGGCCAGCTCCTTCAGCTTCCGCTTCTCCTCAGCCTTCTGGGGGTC
CCATTCCTCTCCACGACGGTAAGAGTCTAGCTCTTCATCTGAGGGTGCAAACT
CCTTTTTGAAGATCATGACATAGCGACAGTCATCATCTTCCCCAAAGGAGAAG
GATGTCAGGCCAGCCACTTCCACCACATCATGTAGTATGCTCCTCTCGATCTT
GTTCATTGGCTGAAACTTTTTCTTGATCTGCCCACTGTCTTGAATGAAATCTG
ACACCTCCTTCTCCATCCTTTTACGAAACTCCACTTTCTGTTGTTTCTCTTGC
TCTTGTAGTTTCTTCAGGCGGGCGGCCTGTTCTCGGGCCTTGCGCCGAGTCTC

CTGGTCACCGAGGCTGGGTGGCTTCTCCATGGAGCTCAGGATGGAGCCCAGTA
GGTCCGCCATCTTGGGAGTC

SEQ ID NO. 30:

PROTEIN

MADLLGSILSSMEKPPSLGDQETRRKAREQAARLKKLQEQEKQQKVEFRKRME
KEVSDFIQDSGQIKKKFQPMNKIERSILHDVVEVAGLTSFSFGEDDDCRYVMI
FKKEFAPSDEELDSYRRGEEWDPQKAEEKRKLKELAQRQEEEAAQQGPVVVSP
ASDYKDKYSHLIGKGAAKDAAHMLQANKTYGCVPVANKRDTRSIEEAMNEIRA
KKRLRQSGEELPPTSRRPAQLPLTPGAGQGAGRDKAAAIRAHPGAPPLNHLLP
AVPQAGGKQVFDLSPLELGYVRGMCVCV

SEQUENCE LISTING

<110> LION Bioscience AG

<120> Novel Cofactors of the Pregnane x Receptor and Methods of Use

<130> L-0017-01-WO-01

<160> 30

<170> PatentIn version 3.0

<210> 1
<211> 675
<212> DNA
<213> Homo sapiens

<400> 1
cttgctctgg ctgtttctgc ccctgggtta acattcaaga tggtacatgc tgaagccttt 60

tctcgtcctt tgagtcggaa tgaagttgtt ggtttaattt tccgtttgac aatatttggt 120

gcagtgacat actttactat caaatggatg gtagatgcaa ttgatccaac cagaaagcaa 180

aaagtagaag ctcagaaaca ggcagaaaaa ctaatgaagc aaattggagt gaaaaatgtg 240

aagctctcaa aatatgaaat gagtattgct gctcatcttg tagaccctct taatatgcat 300

gttacttgga gtgatatagc aggtttagat gatgtcatta cggatctgaa agacacagtc 360

atcttaccta tcaaaaagaa acatttgttt gagaattcca ggcttctgca gcctccaaaa 420

ggtgttcttc tctatgggcc tccaggctgg ggtaaaacgt tgattgccaa ggccacagcc 480

aaagaagcag gctgtccatt tattaacctt cagccttcga cactgaccga taagtggtat 540

ggagaatctc agaaattggc tgctgctgtc ttatcccttg ccataaagct acaaccatcc 600

atcatcttta tagatggaaa tagactcctt ttttacgaaa ccgttcaagt tctgaccatg 660

aaagctacag cccat 675

<210> 2
<211> 675
<212> DNA
<213> Homo sapiens

<400> 2
atgggctgta gctttcatgg tcagaacttg aacggtttcg taaaaaagga gtctatttcc 60

atctataaag atgatggatg gttgtagctt tatggcaagg gataagacag cagcagccaa 120

tttctgagat tctccatacc acttatcggt cagtgtcgaa ggctgaaggt taataaatgg 180

acagcctgct tctttggctg tggccttggc aatcaacgtt ttaccccagc ctggaggccc 240

atagagaaga acacctttg gaggctgcag aagcctggaa ttctcaaaca aatgtttctt 300

tttgataggt aagatgactg tgtctttcag atccgtaatg acatcatcta aacctgctat 360

atcactccaa gtaacatgca tattaagagg gtctacaaga tgagcagcaa tactcatttc 420

atattttgag agcttcacat ttttcactcc aatttgcttc attagttttt ctgcctgttt 480

ctgagcttct acttttttgct ttctggttgg atcaattgca tctaccatcc atttgatagt 540

aaagtatgtc actgcaccaa atattgtcaa acggaaaatt aaaccaacaa cttcattccg 600

actcaaagga cgagaaaagg cttcagcatg taccatcttg aatgttaacc caggggcaga 660

aacagccaga gcaag 675

<210> 3
<211> 225
<212> PRT
<213> Homo sapiens

<400> 3

Leu Ala Leu Ala Val Ser Ala Pro Gly Leu Thr Phe Lys Met Val His
1 5 10 15

Ala Glu Ala Phe Ser Arg Pro Leu Ser Arg Asn Glu Val Val Gly Leu
20 25 30

Ile Phe Arg Leu Thr Ile Phe Gly Ala Val Thr Tyr Phe Thr Ile Lys
35 40 45

Trp Met Val Asp Ala Ile Asp Pro Thr Arg Lys Gln Lys Val Glu Ala
50 55 60

Gln Lys Gln Ala Glu Lys Leu Met Lys Gln Ile Gly Val Lys Asn Val
65 70 75 80

Lys Leu Ser Lys Tyr Glu Met Ser Ile Ala Ala His Leu Val Asp Pro
85 90 95

Leu Asn Met His Val Thr Trp Ser Asp Ile Ala Gly Leu Asp Asp Val
100 105 110

Ile Thr Asp Leu Lys Asp Thr Val Ile Leu Pro Ile Lys Lys Lys His
115 120 125

Leu Phe Glu Asn Ser Arg Leu Leu Gln Pro Pro Lys Gly Val Leu Leu
130 135 140

Tyr Gly Pro Pro Gly Trp Gly Lys Thr Leu Ile Ala Lys Ala Thr Ala
145 150 155 160

Lys Glu Ala Gly Cys Pro Phe Ile Asn Leu Gln Pro Ser Thr Leu Thr
165 170 175

Asp Lys Trp Tyr Gly Glu Ser Gln Lys Leu Ala Ala Ala Val Leu Ser
180 185 190

Leu Ala Ile Lys Leu Gln Pro Ser Ile Ile Phe Ile Asp Gly Asn Arg
195 200 205

Leu Leu Phe Tyr Glu Thr Val Gln Val Leu Thr Met Lys Ala Thr Ala
210 215 220

His
225

<210> 4
<211> 728
<212> DNA
<213> Homo sapiens

<400> 4
gatttagtaa gtcaccatgt gaggacaaaa cttgatgaac tgaaaaggca agaagtagga 60

aggttaagaa tgttaattaa agctaagttg gattcccttc aagatatagg catggaccac 120

caagctcttc taaaacaatt tgatcaccta aaccacctga atcctgacaa gtttgaatcc 180

acagatttag atatgctaat caaagcggca acaagtgatc tggaacacta tgacaagact 240

cgtcatgaag aatttaaaaa atatgaaatg atgaaggaac atgaaaggag agaatattta 300

aaaacattga atgaagaaaa gagaaaagaa gaagagtcta aatttgaaga aatgaagaaa 360

aagcatgaaa atcaccctaa agttaatcac ccaggaagca aagatcaact aaaagaggta 420

tgggaagaga ctgatggatt ggatcctaat gactttgacc ccaagacatt tttcaaatta 480

catgatgtca atagtgatgg attcctggat gaacaagaat tagaagccct atttactaaa 540

gagttggaga aagtttatga ccctaaaaat gaaaaggatg atatggtaga aatggaagaa 600

gaaaggctta aaatgaggga acatgtaatg aatgaggttg atactaacaa agacagattg 660

gtgactcttg gaggagtttt tgaaagccac agaaaaaaaa agaatttttg gagcccagat 720

agctggga 728

<210> 5
<211> 728
<212> DNA
<213> Homo sapiens

<400> 5
tcccagctat ctgggctcca aaaattcttt ttttttctgt ggctttcaaa aactcctcca 60

agagtcacca atctgtcttt gttagtatca acctcattca ttacatgttc cctcatttta 120

agcctttctt cttccatttc taccatatca tccttttcat ttttagggtc ataaactttc 180

tccaactctt tagtaaatag ggcttctaat tcttgttcat ccaggaatcc atcactattg 240

acatcatgta atttgaaaaa tgtcttgggg tcaaagtcat taggatccaa tccatcagtc 300

tcttcccata cctcttttag ttgatctttg cttcctgggt gattaacttt agggtgattt 360

tcatgctttt tcttcatttc ttcaaattta gactcttctt cttttctctt ttcttcattc 420

aatgttttta aatattctct cctttcatgt tccttcatca tttcatattt tttaaattct 480

tcatgacgag tcttgtcata gtgttccaga tcacttgttg ccgctttgat tagcatatct 540

aaatctgtgg attcaaactt gtcaggattc aggtggttta ggtgatcaaa ttgttttaga 600

agagcttggt ggtccatgcc tatatcttga agggaatcca acttagcttt aattaacatt 660

cttaaccttc ctacttcttg ccttttcagt tcatcaagtt ttgtcctcac atggtgactt 720

actaaatc 728

<210> 6
<211> 242
<212> PRT
<213> Homo sapiens

<400> 6

Asp Leu Val Ser His His Val Arg Thr Lys Leu Asp Glu Leu Lys Arg
1 5 10 15

Gln Glu Val Gly Arg Leu Arg Met Leu Ile Lys Ala Lys Leu Asp Ser
20 25 30

Leu Gln Asp Ile Gly Met Asp His Gln Ala Leu Leu Lys Gln Phe Asp
35 40 45

His Leu Asn His Leu Asn Pro Asp Lys Phe Glu Ser Thr Asp Leu Asp
50 55 60

Met Leu Ile Lys Ala Ala Thr Ser Asp Leu Glu His Tyr Asp Lys Thr
65 70 75 80

Arg His Glu Glu Phe Lys Lys Tyr Glu Met Met Lys Glu His Glu Arg
85 90 95

Arg Glu Tyr Leu Lys Thr Leu Asn Glu Glu Lys Arg Lys Glu Glu Glu
100 105 110

Ser Lys Phe Glu Glu Met Lys Lys Lys His Glu Asn His Pro Lys Val
115 120 125

Asn His Pro Gly Ser Lys Asp Gln Leu Lys Glu Val Trp Glu Glu Thr
130 135 140

Asp Gly Leu Asp Pro Asn Asp Phe Asp Pro Lys Thr Phe Phe Lys Leu
145 150 155 160

His Asp Val Asn Ser Asp Gly Phe Leu Asp Glu Gln Glu Leu Glu Ala
165 170 175

Leu Phe Thr Lys Glu Leu Glu Lys Val Tyr Asp Pro Lys Asn Glu Lys
180 185 190

Asp Asp Met Val Glu Met Glu Glu Glu Arg Leu Lys Met Arg Glu His
195 200 205

Val Met Asn Glu Val Asp Thr Asn Lys Asp Arg Leu Val Thr Leu Gly
210 215 220

Gly Val Phe Glu Ser His Arg Lys Lys Lys Asn Phe Trp Ser Pro Asp
225 230 235 240

Ser Trp

<210> 7
<211> 660
<212> DNA
<213> Homo sapiens

<400> 7
tttgtcagac accacgtccg cacaaagctg gatgagctca agcgacagga ggtgtcacgg 60

ctgcggatgc tgctcaaggc caagatggac gccgagcagg atcccaatgt acaggtggat 120

catctgaatc tcctgaaaca gtttgaacac ctggaccctc agaaccagca tacattcgag 180

gcccgcgacc tggagctgct gatccagacg gccacccggg accttgccca gtacgacgca 240

acccatcatg aaaagttcaa gcgctacgag atgcttaagg aacacgagag acggcgttat 300

ctggagtcac tgggagagga gcagagaaag gaggcggaga ggaagctgga agagcaacag 360

cgccggcacc gcgagcaccc taaagtcaac gtgcctggca gccaagccca gttgaaggag 420

gtgtgggagg agctggatgg actggacccc aacaggttta accccaagac cttcttcata 480

ctgcatgata tcaacagtga tggtgtcctg gatgagcagg agctggaggc actcttcacc 540

aaggagctgg agaaagtgta cgacccaaag aatgaggagg acgacatgcg ggagatggag 600

gaggagcgac tgcgcatgct gaagcatgtg atgaagaatg tggacaccca accaggaccg 660

<210> 8
<211> 660
<212> DNA
<213> Homo sapiens

<400> 8
cggtcctggt tgggtgtcca cattcttcat cacatgcttc agcatgcgca gtcgctcctc 60

ctccatctcc cgcatgtcgt cctcctcatt ctttgggtcg tacactttct ccagctcctt 120

ggtgaagagt gcctccagct cctgctcatc caggacacca tcactgttga tatcatgcag 180

tatgaagaag gtcttggggt taaacctgtt ggggtccagt ccatccagct cctcccacac 240

ctccttcaac tgggcttggc tgccaggcac gttgacttta gggtgctcgc ggtgccggcg 300

ctgttgctct tccagcttcc tctccgcctc ctttctctgc tcctctccca gtgactccag 360

ataacgccgt ctctcgtgtt ccttaagcat ctcgtagcgc ttgaactttt catgatgggt 420

tgcgtcgtac tgggcaaggt cccgggtggc cgtctggatc agcagctcca ggtcgcgggc 480

ctcgaatgta tgctggttct gagggtccag gtgttcaaac tgtttcagga gattcagatg 540

atccacctgt acattgggat cctgctcggc gtccatcttg gccttgagca gcatccgcag 600

ccgtgacacc tcctgtcgct tgagctcatc cagctttgtg cggacgtggt gtctgacaaa 660

<210> 9
<211> 220
<212> PRT
<213> Homo sapiens

<400> 9

Phe Val Arg His His Val Arg Thr Lys Leu Asp Glu Leu Lys Arg Gln
1 5 10 15

Glu Val Ser Arg Leu Arg Met Leu Leu Lys Ala Lys Met Asp Ala Glu
20 25 30

Gln Asp Pro Asn Val Gln Val Asp His Leu Asn Leu Leu Lys Gln Phe
35 40 45

Glu His Leu Asp Pro Gln Asn Gln His Thr Phe Glu Ala Arg Asp Leu
50 55 60

Glu Leu Leu Ile Gln Thr Ala Thr Arg Asp Leu Ala Gln Tyr Asp Ala
65 70 75 80

Thr His His Glu Lys Phe Lys Arg Tyr Glu Met Leu Lys Glu His Glu
85 90 95

Arg Arg Arg Tyr Leu Glu Ser Leu Gly Glu Glu Gln Arg Lys Glu Ala
100 105 110

Glu Arg Lys Leu Glu Glu Gln Gln Arg Arg His Arg Glu His Pro Lys
115 120 125

Val Asn Val Pro Gly Ser Gln Ala Gln Leu Lys Glu Val Trp Glu Glu
130 135 140

Leu Asp Gly Leu Asp Pro Asn Arg Phe Asn Pro Lys Thr Phe Phe Ile
145 150 155 160

Leu His Asp Ile Asn Ser Asp Gly Val Leu Asp Glu Gln Glu Leu Glu
165 170 175

Ala Leu Phe Thr Lys Glu Leu Glu Lys Val Tyr Asp Pro Lys Asn Glu
180 185 190

Glu Asp Asp Met Arg Glu Met Glu Glu Glu Arg Leu Arg Met Leu Lys
195 200 205

His Val Met Lys Asn Val Asp Thr Gln Pro Gly Pro
210 215 220

<210> 10
<211> 670
<212> DNA
<213> Homo sapiens

<400> 10
ggggactcgg ccctgaacga gcaggagaag gagttgcagc ggcggctgaa gcgtctctac 60

ccggccgagg acgaacaaga gacgccgctg cctaggtcct ggagcccgaa ggacaagttc 120

agctacatcg gcctctctca gaacaacctg cgggtgcact acaaaggtca tggcaaaacc 180

ccaaaagatg ccgcgtcagt tcgagccacg catccaatac cagcagcctg tgggatttat 240

tattttgaag taaaaattgt cagtaaggga agagatggtt acatgggaat tggtctttct 300

gctcaaggtg tgaacatgaa tagactacca ggttgggata agcattcata tggttaccat 360

ggggatgatg gacattcgtt ttgttcttct ggaactggac aaccttatgg accaactttc 420

actactggtg atgtcattgg ctgttgtgtt aatcttatca acaatacctg cttttacacc 480

aagaatggac atagtttagg tattgcttta ctgacctacc gccaaatttg tatcctactg 540

tggggcttca aacaccagga gaagtggtcg atgccaattt ttgggcaaca tcctttccgt 600

gtttgatata aaaaactata tgccgggagt ggagaaccaa aatccaggcc ccagatagat 660

ccgatttcct 670

<210> 11
<211> 670
<212> DNA
<213> Homo sapiens

<400> 11
aggaaatcgg atctatctgg ggcctggatt ttggttctcc actcccggca tatagttttt 60

tatatcaaac acggaaagga tgttgcccaa aaattggcat cgaccacttc tcctggtgtt 120

tgaagcccca cagtaggata caaatttggc ggtaggtcag taaagcaata cctaaactat 180

gtccattctt ggtgtaaaag caggtattgt tgataagatt aacacaacag ccaatgacat 240

caccagtagt gaaagttggt ccataaggtt gtccagttcc agaagaacaa aacgaatgtc 300

catcatcccc atggtaacca tatgaatgct tatcccaacc tggtagtcta ttcatgttca 360

caccttgagc agaaagacca attcccatgt aaccatctct tcccttactg acaattttta 420

cttcaaaata ataaatccca caggctgctg gtattggatg cgtggctcga actgacgcgg 480

catcttttgg ggttttgcca tgacctttgt agtgcacccg caggttgttc tgagagaggc 540

cgatgtagct gaacttgtcc ttcgggctcc aggacctagg cagcggcgtc tcttgttcgt 600

cctcggccgg gtagagacgc ttcagccgcc gctgcaactc cttctcctgc tcgttcaggg 660

ccgagtcccc 670

<210> 12
<211> 201
<212> PRT
<213> Homo sapiens

<400> 12

Gly Asp Ser Ala Leu Asn Glu Gln Glu Lys Glu Leu Gln Arg Arg Leu
1 5 10 15

Lys Arg Leu Tyr Pro Ala Glu Asp Glu Gln Glu Thr Pro Leu Pro Arg
20 25 30

Ser Trp Ser Pro Lys Asp Lys Phe Ser Tyr Ile Gly Leu Ser Gln Asn
35 40 45

Asn Leu Arg Val His Tyr Lys Gly His Gly Lys Thr Pro Lys Asp Ala
50 55 60

Ala Ser Val Arg Ala Thr His Pro Ile Pro Ala Ala Cys Gly Ile Tyr
65 70 75 80

Tyr Phe Glu Val Lys Ile Val Ser Lys Gly Arg Asp Gly Tyr Met Gly
85 90 95

Ile Gly Leu Ser Ala Gln Gly Val Asn Met Asn Arg Leu Pro Gly Trp
100 105 110

Asp Lys His Ser Tyr Gly Tyr His Gly Asp Asp Gly His Ser Phe Cys
115 120 125

Ser Ser Gly Thr Gly Gln Pro Tyr Gly Pro Thr Phe Thr Thr Gly Asp
130 135 140

Val Ile Gly Cys Cys Val Asn Leu Ile Asn Asn Thr Cys Phe Tyr Thr
145 150 155 160

Lys Asn Gly His Ser Leu Gly Ile Ala Leu Leu Thr Tyr Arg Gln Ile
165 170 175

Cys Ile Leu Leu Trp Gly Phe Lys His Gln Glu Lys Trp Ser Met Pro
180 185 190

Ile Phe Gly Gln His Pro Phe Arg Val
195 200

<210> 13
<211> 1140
<212> DNA
<213> Homo sapiens

<400> 13
atgacatgtg aaggatgcaa gggctttttc aggagggcca tgaaacgcaa cgcccggctg 60

aggtgcccct tccggaaggg cgcctgcgag atcacccgga agacccggcg acagtgccag 120

gcctgccgcc tgcgcaagtg cctggagagc ggcatgaaga aggagatgat catgtccgac 180

gaggccgtgg aggagaggcg ggccttgatc aagcggaaga aaagtgaacg gacagggact 240

cagccactgg gagtgcaggg gctgacagag gagcagcgga tgatgatcag ggagctgatg 300

gacgctcaga tgaaaacctt tgacactacc ttctcccatt tcaagaattt ccggctgcca 360

ggggtgctta gcagtggctg cgagttgcca gagtctctgc aggccccatc gagggaagaa 420

gctgccaagt ggagccaggt ccggaaagat ctgtgctctt tgaaggtctc tctgcagctg 480

cgggggggagg atggcagtgt ctggaactac aaacccccag ccgacagtgg cgggaaagag 540

atcttctccc tgctgcccca catggctgac atgtcaacct acatgttcaa aggcatcatc 600

agctttgcca aagtcatctc ctacttcagg gacttgccca tcgaggacca gatctccctg 660

ctgaaggggg ccgctttcga gctgtgtcaa ctgagattca acacagtgtt caacgcggag 720

actggaacct gggagtgtgg ccggctgtcc tactgcttgg aagacactgc aggtggcttc 780

cagcaacttc tactggagcc catgctgaaa ttccactaca tgctgaagaa gctgcagctg 840

catgaggagg agtatgtgct gatgcaggcc atctccctct tctccccaga ccgcccaggt 900

gtgctgcagc accgcgtggt ggaccagctg caggagcaat tcgccattac tctgaagtcc 960

tacattgaat gcaatcggcc ccagcctgct cataggttct tgttcctgaa gatcatggct 1020

atgctcaccg agctccgcag catcaatgct cagcacaccc agcggctgct gcgcatccag 1080

gacatacacc cctttgctac gcccctcatg caggagttgt tcggcatcac aggtagctga 1140

<210> 14
<211> 1140
<212> DNA
<213> Homo sapiens

<400> 14
agtcgatgga cactacggct tgttgaggac gtactccccg catcgtttcc ccacatacag 60

gacctacgcg tcgtcggcga cccacacgac tcgtaactac gacgcctcga gccactcgta 120

tcggtactag aagtccttgt tcttggatac tcgtccgacc ccggctaacg taagttacat 180

cctgaagtct cattaccgct taacgaggac gtcgaccagg tggtgcgcca cgacgtcgtg 240

tggacccgcc agacccctct tctccctcta ccggacgtag tcgtgtatga ggaggagtac 300

gtcgacgtcg aagaagtcgt acatcacctt aaagtcgtac ccgaggtcat cttcaacgac 360

cttcggtgga cgtcacagaa ggttcgtcat cctgtcggcc ggtgtgaggg tccaaggtca 420

gaggcgcaac ttgtgacaca acttagagtc aactgtgtcg agctttcgcc gggggaagtc 480

gtccctctag accaggagct acccgttcag ggacttcatc ctctactgaa accgtttcga 540

ctactacgga aacttgtaca tccaactgta cagtcggtac accccgtcgt ccctcttcta 600

gagaaagggc ggtgacagcc gacccccaaa catcaaggtc tgtgacggta ggaggggggc 660

gtcgacgtct ctctggaagt ttctcgtgtc tagaaaggcc tggaccgagg tgaaccgtcg 720

aagaagggag ctaccccgga cgtctctgag accgttgagc gtcggtgacg attcgtgggg 780

accgtcggcc tttaagaact ttaccctctt ccatcacagt ttccaaaagt agactcgcag 840

gtagtcgagg gactagtagt aggcgacgag gagacagtcg gggacgtgag ggtcaccgac 900

tcagggacag gcaagtgaaa agaaggcgaa ctagttccgg gcggagagga ggtgccggag 960

cagcctgtac tagtagagga agaagtacgg cgagaggtcc gtgaacgcgt ccgccgtccg 1020

gaccgtgaca gcggcccaga aggcccacta gagcgtccgc gggaaggcct tccccgtgga 1080

gtcggcccgc aacgcaaagt accgggagga ctttttcggg aacgtaggaa gtgtacagta 1140

<210> 15
<211> 434
<212> PRT
<213> Homo sapiens

<400> 15

Met Glu Val Arg Pro Lys Glu Ser Trp Asn His Ala Asp Phe Val His
1 5 10 15

Cys Glu Asp Thr Glu Ser Val Pro Gly Lys Pro Ser Val Asn Ala Asp
20 25 30

Glu Glu Val Gly Gly Pro Gln Ile Cys Arg Val Cys Gly Asp Lys Ala
35 40 45

```
Thr Gly Tyr His Phe Asn Val Met Thr Cys Glu Gly Cys Lys Gly Phe
    50                  55                  60

Phe Arg Arg Ala Met Lys Arg Asn Ala Arg Leu Arg Cys Pro Phe Arg
65                  70                  75                  80

Lys Gly Ala Cys Glu Ile Thr Arg Lys Thr Arg Arg Gln Cys Gln Ala
                85                  90                  95

Cys Arg Leu Arg Lys Cys Leu Glu Ser Gly Met Lys Lys Glu Met Ile
            100                 105                 110

Met Ser Asp Glu Ala Val Glu Glu Arg Arg Ala Leu Ile Lys Arg Lys
        115                 120                 125

Lys Ser Glu Arg Thr Gly Thr Gln Pro Leu Gly Val Gln Gly Leu Thr
    130                 135                 140

Glu Glu Gln Arg Met Met Ile Arg Glu Leu Met Asp Ala Gln Met Lys
145                 150                 155                 160

Thr Phe Asp Thr Thr Phe Ser His Phe Lys Asn Phe Arg Leu Pro Gly
                165                 170                 175

Val Leu Ser Ser Gly Cys Glu Leu Pro Glu Ser Leu Gln Ala Pro Ser
            180                 185                 190

Arg Glu Glu Ala Ala Lys Trp Ser Gln Val Arg Lys Asp Leu Cys Ser
        195                 200                 205

Leu Lys Val Ser Leu Gln Leu Arg Gly Glu Asp Gly Ser Val Trp Asn
    210                 215                 220

Tyr Lys Pro Pro Ala Asp Ser Gly Gly Lys Glu Ile Phe Ser Leu Leu
225                 230                 235                 240

Pro His Met Ala Asp Met Ser Thr Tyr Met Phe Lys Gly Ile Ile Ser
                245                 250                 255

Phe Ala Lys Val Ile Ser Tyr Phe Arg Asp Leu Pro Ile Glu Asp Gln
            260                 265                 270

Ile Ser Leu Leu Lys Gly Ala Ala Phe Glu Leu Cys Gln Leu Arg Phe
        275                 280                 285

Asn Thr Val Phe Asn Ala Glu Thr Gly Thr Trp Glu Cys Gly Arg Leu
    290                 295                 300

Ser Tyr Cys Leu Glu Asp Thr Ala Gly Gly Phe Gln Gln Leu Leu Leu
305                 310                 315                 320

Glu Pro Met Leu Lys Phe His Tyr Met Leu Lys Lys Leu Gln Leu His
                325                 330                 335

Glu Glu Glu Tyr Val Leu Met Gln Ala Ile Ser Leu Phe Ser Pro Asp
            340                 345                 350

Arg Pro Gly Val Leu Gln His Arg Val Val Asp Gln Leu Gln Glu Gln
        355                 360                 365

Phe Ala Ile Thr Leu Lys Ser Tyr Ile Glu Cys Asn Arg Pro Gln Pro
    370                 375                 380
```

Ala His Arg Phe Leu Phe Leu Lys Ile Met Ala Met Leu Thr Glu Leu
385 390 395 400

Arg Ser Ile Asn Ala Gln His Thr Gln Arg Leu Leu Arg Ile Gln Asp
405 410 415

Ile His Pro Phe Ala Thr Pro Leu Met Gln Glu Leu Phe Gly Ile Thr
420 425 430

Gly Ser

<210> 16
<211> 990
<212> DNA
<213> Homo sapiens

<400> 16
ggcatgaaga aggagatgat catgtccgac gaggccgtgg aggagaggcg ggccttgatc 60

aagcggaaga aaagtgaacg gacagggact cagccactgg gagtgcaggg gctgacagag 120

gagcagcgga tgatgatcag ggagctgatg gacgctcaga tgaaaacctt tgacactacc 180

ttctcccatt tcaagaattt ccggctgcca ggggtgctta gcagtggctg cgagttgcca 240

gagtctctgc aggccccatc gagggaagaa gctgccaagt ggagccaggt ccggaaagat 300

ctgtgctctt tgaaggtctc tctgcagctg cgggggggagg atggcagtgt ctggaactac 360

aaaccccccag ccgacagtgg cgggaaagag atcttctccc tgctgcccca catggctgac 420

atgtcaacct acatgttcaa aggcatcatc agctttgcca aagtcatctc ctacttcagg 480

gacttgccca tcgaggacca gatctccctg ctgaaggggg ccgctttcga gctgtgtcaa 540

ctgagattca acacagtgtt caacgcggag actggaacct gggagtgtgg ccggctgtcc 600

tactgcttgg aagacactgc aggtggcttc cagcaacttc tactggagcc catgctgaaa 660

ttccactaca tgctgaagaa gctgcagctg catgaggagg agtatgtgct gatgcaggcc 720

atctccctct tctccccaga ccgcccaggt gtgctgcagc accgcgtggt ggaccagctg 780

caggagcaat tcgccattac tctgaagtcc tacattgaat gcaatcggcc ccagcctgct 840

cataggttct tgttcctgaa gatcatggct atgctcaccg agctccgcag catcaatgct 900

cagcacaccc agcggctgct gcgcatccag gacatacacc cctttgctac gcccctcatg 960

caggagttgt tcggcatcac aggtagctga 990

<210> 17
<211> 990
<212> DNA
<213> Homo sapiens

<400> 17
tcagctacct gtgatgccga acaactcctg catgaggggc gtagcaaagg ggtgtatgtc 60

ctggatgcgc agcagccgct gggtgtgctg agcattgatg ctgcggagct cggtgagcat 120

agccatgatc ttcaggaaca agaacctatg agcaggctgg ggccgattgc attcaatgta 180

ggacttcaga gtaatggcga attgctcctg cagctggtcc accacgcggt gctgcagcac 240

acctgggcgg tctggggaga agagggagat ggcctgcatc agcacatact cctcctcatg 300

cagctgcagc ttcttcagca tgtagtggaa tttcagcatg ggctccagta gaagttgctg 360

gaagccacct gcagtgtctt ccaagcagta ggacagccgg ccacactccc aggttccagt 420

ctccgcgttg aacactgtgt tgaatctcag ttgacacagc tcgaaagcgg ccccccttcag 480

cagggagatc tggtcctcga tgggcaagtc cctgaagtag gagatgactt tggcaaagct 540

gatgatgcct ttgaacatgt aggttgacat gtcagccatg tggggcagca gggagaagat 600

ctctttcccg ccactgtcgg ctgggggttt gtagttccag acactgccat cctccccccg 660

cagctgcaga gagaccttca aagagcacag atctttccgg acctggctcc acttggcagc 720

ttcttccctc gatggggcct gcagagactc tggcaactcg cagccactgc taagcacccc 780

tggcagccgg aaattcttga atgggagaa ggtagtgtca aaggttttca tctgagcgtc 840

catcagctcc ctgatcatca tccgctgctc ctctgtcagc ccctgcactc ccagtggctg 900

agtccctgtc cgttcacttt tcttccgctt gatcaaggcc cgcctctcct ccacggcctc 960

gtcggacatg atcatctcct tcttcatgcc 990

<210> 18
<211> 329
<212> PRT
<213> Homo sapiens

<400> 18

Gly Met Lys Lys Glu Met Ile Met Ser Asp Glu Ala Val Glu Glu Arg
1 5 10 15

Arg Ala Leu Ile Lys Arg Lys Lys Ser Glu Arg Thr Gly Thr Gln Pro
20 25 30

Leu Gly Val Gln Gly Leu Thr Glu Glu Gln Arg Met Met Ile Arg Glu
35 40 45

Leu Met Asp Ala Gln Met Lys Thr Phe Asp Thr Thr Phe Ser His Phe
50 55 60

Lys Asn Phe Arg Leu Pro Gly Val Leu Ser Ser Gly Cys Glu Leu Pro
65 70 75 80

Glu Ser Leu Gln Ala Pro Ser Arg Glu Glu Ala Ala Lys Trp Ser Gln
85 90 95

Val Arg Lys Asp Leu Cys Ser Leu Lys Val Ser Leu Gln Leu Arg Gly
100 105 110

Glu Asp Gly Ser Val Trp Asn Tyr Lys Pro Pro Ala Asp Ser Gly Gly
115 120 125

```
Lys Glu Ile Phe Ser Leu Leu Pro His Met Ala Asp Met Ser Thr Tyr
    130                 135                 140

Met Phe Lys Gly Ile Ile Ser Phe Ala Lys Val Ile Ser Tyr Phe Arg
145                 150                 155                 160

Asp Leu Pro Ile Glu Asp Gln Ile Ser Leu Leu Lys Gly Ala Ala Phe
                165                 170                 175

Glu Leu Cys Gln Leu Arg Phe Asn Thr Val Phe Asn Ala Glu Thr Gly
            180                 185                 190

Thr Trp Glu Cys Gly Arg Leu Ser Tyr Cys Leu Glu Asp Thr Ala Gly
        195                 200                 205

Gly Phe Gln Gln Leu Leu Leu Glu Pro Met Leu Lys Phe His Tyr Met
    210                 215                 220

Leu Lys Lys Leu Gln Leu His Glu Glu Glu Tyr Val Leu Met Gln Ala
225                 230                 235                 240

Ile Ser Leu Phe Ser Pro Asp Arg Pro Gly Val Leu Gln His Arg Val
                245                 250                 255

Val Asp Gln Leu Gln Glu Gln Phe Ala Ile Thr Leu Lys Ser Tyr Ile
            260                 265                 270

Glu Cys Asn Arg Pro Gln Pro Ala His Arg Phe Leu Phe Leu Lys Ile
        275                 280                 285

Met Ala Met Leu Thr Glu Leu Arg Ser Ile Asn Ala Gln His Thr Gln
    290                 295                 300

Arg Leu Leu Arg Ile Gln Asp Ile His Pro Phe Ala Thr Pro Leu Met
305                 310                 315                 320

Gln Glu Leu Phe Gly Ile Thr Gly Ser
                325
```

<210> 19
<211> 1389
<212> DNA
<213> Homo sapiens

<400> 19

```
atggacacca aacatttcct gccgctcgat ttctccaccc aggtgaactc ctccctcacc     60

tccccgacgg ggcgaggctc catggctgcc ccctcgctgc acccgtccct ggggcctggc    120

atcggctccc cgggacagct gcattctccc atcagcaccc tgagctcccc catcaacggc    180

atgggcccgc ctttctcggt catcagctcc cccatgggcc cccactccat gtcggtgccc    240

accacaccca ccctgggctt cagcactggc agcccccagc tcagctcacc tatgaacccc    300

gtcagcagca gcgaggacat caagcccccc ctgggcctca atggcgtcct caaggtcccc    360

gcccacccct caggaaacat ggcttccttc accaagcaca tctgcgccat ctgcggggac    420

cgctcctcag gcaagcacta tggagtgtac agctgcgagg ggtgcaaggg cttcttcaag    480

cggacggtgc gcaaggacct gacctacacc tgccgcgaca acaaggactg cctgattgac    540
```

aagcggcagc ggaaccggtg ccagtactgc cgctaccaga agtgcctggc catgggcatg 600

aagcgggaag ccgtgcagga ggagcggcag cgtggcaagg accggaacga gaatgaggtg 660

gagtcgacca gcagcgccaa cgaggacatg ccggtggaga ggatcctgga ggctgagctg 720

gccgtggagc ccaagaccga gacctacgtg gaggcaaaca tggggctgaa ccccagctcg 780

ccgaacgacc ctgtcaccaa catttgccaa gcagccgaca aacagctttt caccctggtg 840

gagtgggcca agcggatccc acacttctca gagctgcccc tggacgacca ggtcatcctg 900

ctgcgggcag gctggaatga gctgctcatc gcctccttct cccaccgctc catcgccgtg 960

aaggacggga tcctcctggc caccgggctg cacgtccacc ggaacagcgc ccacagcgca 1020

ggggtgggcg ccatctttga cagggtgctg acggagcttg tgtccaagat gcgggacatg 1080

cagatggaca agacggagct gggctgcctg cgcgccatcg tcctctttaa ccctgactcc 1140

aaggggctct cgaacccggc cgaggtggag gcgctgaggg agaaggtcta tgcgtccttg 1200

gaggcctact gcaagcacaa gtacccagag cagccgggaa ggttcgctaa gctcttgctc 1260

cgcctgccgg ctctgcgctc catcgggctc aaatgcctgg aacatctctt cttcttcaag 1320

ctcatcgggg acacacccat tgacaccttc cttatggaga tgctggaggc gccgcaccaa 1380

atgacttag 1389

<210> 20
<211> 1389
<212> DNA
<213> Homo sapiens

<400> 20
ctaagtcatt tggtgcggcg cctccagcat ctccataagg aaggtgtcaa tgggtgtgtc 60

cccgatgagc ttgaagaaga agagatgttc caggcatttg agcccgatgg agcgcagagc 120

cggcaggcgg agcaagagct tagcgaacct tcccggctgc tctgggtact tgtgcttgca 180

gtaggcctcc aaggacgcat agaccttctc cctcagcgcc tccacctcgg ccgggttcga 240

gagccccttg gagtcagggt taaagaggac gatggcgcgc aggcagccca gctccgtctt 300

gtccatctgc atgtcccgca tcttggacac aagctccgtc agcaccctgt caaagatggc 360

gcccaccccc gcgctgtggg cgctgttccg gtggacgtgc agcccggtgg ccaggaggat 420

cccgtccttc acggcgatgg agcggtggga gaaggaggcg atgagcagct cattccagcc 480

tgcccgcagc aggatgacct ggtcgtccag gggcagctct gagaagtgtg ggatccgctt 540

ggcccactcc accagggtga aaagctgttt gtcggctgct tggcaaatgt tggtgacagg 600

gtcgttcggc gagctggggt tcagccccat gtttgcctcc acgtaggtct cggtcttggg 660

ctccacggcc agctcagcct ccaggatcct ctccaccggc atgtcctcgt tggcgctgct 720

ggtcgactcc acctcattct cgttccggtc cttgccacgc tgccgctcct cctgcacggc 780

ttcccgcttc atgcccatgg ccaggcactt ctggtagcgg cagtactggc accggttccg 840

ctgccgcttg tcaatcaggc agtccttgtt gtcgcggcag gtgtaggtca ggtccttgcg 900

caccgtccgc ttgaagaagc ccttgcaccc ctcgcagctg tacactccat agtgcttgcc 960

tgaggagcgg tccccgcaga tggcgcagat gtgcttggtg aaggaagcca tgtttcctga 1020

ggggtgggcg gggaccttga ggacgccatt gaggcccagg gggggcttga tgtcctcgct 1080

gctgctgacg gggttcatag gtgagctgag ctgggggctg ccagtgctga agcccagggt 1140

gggtgtggtg ggcaccgaca tggagtgggg gcccatgggg gagctgatga ccgagaaagg 1200

cgggcccatg ccgttgatgg gggagctcag ggtgctgatg ggagaatgca gctgtcccgg 1260

ggagccgatg ccaggcccca gggacgggtg cagcgagggg gcagccatgg agcctcgccc 1320

cgtcggggag gtgagggagg agttcacctg ggtggagaaa tcgagcggca ggaaatgttt 1380

ggtgtccat 1389

<210> 21
<211> 462
<212> PRT
<213> Homo sapiens

<400> 21

Met Asp Thr Lys His Phe Leu Pro Leu Asp Phe Ser Thr Gln Val Asn
1 5 10 15

Ser Ser Leu Thr Ser Pro Thr Gly Arg Gly Ser Met Ala Ala Pro Ser
20 25 30

Leu His Pro Ser Leu Gly Pro Gly Ile Gly Ser Pro Gly Gln Leu His
35 40 45

Ser Pro Ile Ser Thr Leu Ser Ser Pro Ile Asn Gly Met Gly Pro Pro
50 55 60

Phe Ser Val Ile Ser Ser Pro Met Gly Pro His Ser Met Ser Val Pro
65 70 75 80

Thr Thr Pro Thr Leu Gly Phe Ser Thr Gly Ser Pro Gln Leu Ser Ser
85 90 95

Pro Met Asn Pro Val Ser Ser Ser Glu Asp Ile Lys Pro Pro Leu Gly
100 105 110

Leu Asn Gly Val Leu Lys Val Pro Ala His Pro Ser Gly Asn Met Ala
115 120 125

Ser Phe Thr Lys His Ile Cys Ala Ile Cys Gly Asp Arg Ser Ser Gly
130 135 140

Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys
145 150 155 160

Arg Thr Val Arg Lys Asp Leu Thr Tyr Thr Cys Arg Asp Asn Lys Asp
165 170 175

Cys Leu Ile Asp Lys Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr
180 185 190

Gln Lys Cys Leu Ala Met Gly Met Lys Arg Glu Ala Val Gln Glu Glu
195 200 205

Arg Gln Arg Gly Lys Asp Arg Asn Glu Asn Glu Val Glu Ser Thr Ser
210 215 220

Ser Ala Asn Glu Asp Met Pro Val Glu Arg Ile Leu Glu Ala Glu Leu
225 230 235 240

Ala Val Glu Pro Lys Thr Glu Thr Tyr Val Glu Ala Asn Met Gly Leu
245 250 255

Asn Pro Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala Ala
260 265 270

Asp Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His
275 280 285

Phe Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly
290 295 300

Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val
305 310 315 320

Lys Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser
325 330 335

Ala His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu
340 345 350

Leu Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly
355 360 365

Cys Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser
370 375 380

Asn Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu
385 390 395 400

Glu Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala
405 410 415

Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys
420 425 430

Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp
435 440 445

Thr Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Met Thr
450 455 460

<210> 22
<211> 1602
<212> DNA
<213> Homo sapiens

<400> 22
atgtcttggg ccgctcgccc gcccttcctc cctcagcggc atgccgcagg gcagtgtggg 60

ccggtggggg tgcgaaaaga aatgcattgt ggggtcgcgt cccggtggcg gcggcgacgg 120

ccctggctgg atcccgcagc ggcggcggcg gcggcggtgg caggcggaga acaacaaacc 180

ccggagccgg agccagggga ggctggacgg gacgggatgg gcgacagcgg gcgggactcc 240

cgaagcccag acagctcctc cccaaatccc cttccccagg gagtccctcc cccttctcct 300

cctgggccac ccctaccccc ttcaacagct cctacccttg gaggctctgg ggccccaccc 360

ccacccccga tgccaccacc cccactgggc tctcccttc cagtcatcag ttcttccatg 420

gggtcccctg gtctgccccc tccagctccc ccaggattct ccgggcctgt cagcagcccc 480

cagattaact caacagtgtc actccctggg ggtgggtctg gccccctga agatgtgaag 540

ccaccagtct taggggtccg gggcctgcac tgtccacccc ctccaggtgg ccctggggct 600

ggcaaacggc tatgtgcaat ctgcggggac agaagctcag gcaaacacta cggggtttac 660

agctgtgagg gttgcaaggg cttcttcaaa cgcaccatcc gcaaagacct tacatactct 720

tgccgggaca acaaagactg cacagtggac aagcgccagc ggaaccgctg tcagtactgc 780

cgctatcaga agtgcctggc cactggcatg aagagggagg cggtacagga ggagcgtcag 840

cggggaaagg acaaggatgg ggatggggag ggggctgggg gagcccccga ggagatgcct 900

gtggacagga tcctggaggc agagcttgct gtggaacaga agagtgacca gggcgttgag 960

ggtcctgggg gaaccggggg tagcggcagc agcccaaatg accctgtgac taacatctgt 1020

caggcagctg acaaacagct attcacgctt gttgagtggg cgaagaggat cccacacttt 1080

tcctccttgc ctctggatga tcaggtcata ttgctgcggg caggctggaa tgaactcctc 1140

attgcctcct tttcacaccg atccattgat gttcgagatg gcatcctcct tgccacaggt 1200

cttcacgtgc accgcaactc agcccattca gcaggagtag gagccatctt tgatcgggtg 1260

ctgacagagc tagtgtccaa aatgcgtgac atgaggatgg acaagacaga gcttggctgc 1320

ctgagggcaa tcattctgtt taatccagat gccaagggcc tctccaaccc tagtgaggtg 1380

gaggtcctgc gggagaaagt gtatgcatca ctggagacct actgcaaaca gaagtaccct 1440

gagcagcagg gacggtttgc caagctgctg ctacgtcttc ctgccctccg gtccattggc 1500

cttaagtgtc tagagcatct gtttttcttc aagctcattg gtgacacccc catcgacacc 1560

ttcctcatgg agatgcttga ggctccccat caactggcct ga 1602

<210> 23
<211> 1602
<212> DNA
<213> Homo sapiens

<400> 23
tcaggccagt tgatggggag cctcaagcat ctccatgagg aaggtgtcga tgggggtgtc 60

accaatgagc ttgaagaaaa acagatgctc tagacactta aggccaatgg accggagggc 120

aggaagacgt agcagcagct tggcaaaccg tccctgctgc tcagggtact tctgtttgca 180

gtaggtctcc agtgatgcat acactttctc ccgcaggacc tccacctcac tagggttgga 240

gaggcccttg gcatctggat taaacagaat gattgccctc aggcagccaa gctctgtctt 300

gtccatcctc atgtcacgca ttttggacac tagctctgtc agcacccgat caaagatggc 360

tcctactcct gctgaatggg ctgagttgcg gtgcacgtga agacctgtgg caaggaggat 420

gccatctcga acatcaatgg atcggtgtga aaaggaggca atgaggagtt cattccagcc 480

tgcccgcagc aatatgacct gatcatccag aggcaaggag gaaaagtgtg ggatcctctt 540

cgcccactca acaagcgtga atagctgttt gtcagctgcc tgacagatgt tagtcacagg 600

gtcatttggg ctgctgccgc tacccccggt tcccccagga ccctcaacgc cctggtcact 660

cttctgttcc acagcaagct ctgcctccag gatcctgtcc acaggcatct cctcgggggc 720

tcccccagcc ccctccccat ccccatcctt gtcctttccc cgctgacgct cctcctgtac 780

cgcctccctc ttcatgccag tggccaggca cttctgatag cggcagtact gacagcggtt 840

ccgctggcgc ttgtccactg tgcagtcttt gttgtcccgg caagagtatg taaggtcttt 900

gcggatggtg cgtttgaaga agcccttgca accctcacag ctgtaaaccc cgtagtgttt 960

gcctgagctt ctgtccccgc agattgcaca tagccgtttg ccagccccag ggccacctgg 1020

agggggtgga cagtgcaggc cccggacccc taagactggt ggcttcacat cttcaggggg 1080

gccagaccca cccccaggga gtgacactgt tgagttaatc tgggggctgc tgacaggccc 1140

ggagaatcct gggggagctg gaggggggcag accaggggac cccatggaag aactgatgac 1200

tggaaaggga gagcccagtg ggggtggtgg catcgggggt gggggtgggg ccccagagcc 1260

tccaagggta ggagctgttg aagggggtag gggtggccca ggaggagaag ggggagggac 1320

tccctgggga aggggatttg gggaggagct gtctgggctt cgggagtccc gcccgctgtc 1380

gcccatcccg tcccgtccag cctcccctgg ctccggctcc ggggtttgtt gttctccgcc 1440

tgccaccgcc gccgccgccg ccgctgcggg atccagccag ggccgtcgcc gccgccaccg 1500

ggacgcgacc ccacaatgca tttcttttcg caccccccacc ggcccacact gccctgcggc 1560

atgccgctga gggaggaagg gcgggcgagc ggcccaagac at 1602

<210> 24
<211> 533
<212> PRT
<213> Homo sapiens

<400> 24

Met Ser Trp Ala Ala Arg Pro Pro Phe Leu Pro Gln Arg His Ala Ala
1 5 10 15

Gly Gln Cys Gly Pro Val Gly Val Arg Lys Glu Met His Cys Gly Val
20 25 30

```
Ala Ser Arg Trp Arg Arg Arg Arg Pro Trp Leu Asp Pro Ala Ala Ala
        35                  40                  45

Ala Ala Ala Ala Val Ala Gly Gly Glu Gln Gln Thr Pro Glu Pro Glu
    50                  55                  60

Pro Gly Glu Ala Gly Arg Asp Gly Met Gly Asp Ser Gly Arg Asp Ser
65                  70                  75                  80

Arg Ser Pro Asp Ser Ser Ser Pro Asn Pro Leu Pro Gln Gly Val Pro
                85                  90                  95

Pro Pro Ser Pro Pro Gly Pro Pro Leu Pro Pro Ser Thr Ala Pro Thr
            100                 105                 110

Leu Gly Gly Ser Gly Ala Pro Pro Pro Pro Pro Met Pro Pro Pro Pro
        115                 120                 125

Leu Gly Ser Pro Phe Pro Val Ile Ser Ser Ser Met Gly Ser Pro Gly
    130                 135                 140

Leu Pro Pro Pro Ala Pro Pro Gly Phe Ser Gly Pro Val Ser Ser Pro
145                 150                 155                 160

Gln Ile Asn Ser Thr Val Ser Leu Pro Gly Gly Gly Ser Gly Pro Pro
                165                 170                 175

Glu Asp Val Lys Pro Pro Val Leu Gly Val Arg Gly Leu His Cys Pro
            180                 185                 190

Pro Pro Pro Gly Gly Pro Gly Ala Gly Lys Arg Leu Cys Ala Ile Cys
        195                 200                 205

Gly Asp Arg Ser Ser Gly Lys His Tyr Gly Val Tyr Ser Cys Glu Gly
    210                 215                 220

Cys Lys Gly Phe Phe Lys Arg Thr Ile Arg Lys Asp Leu Thr Tyr Ser
225                 230                 235                 240

Cys Arg Asp Asn Lys Asp Cys Thr Val Asp Lys Arg Gln Arg Asn Arg
                245                 250                 255

Cys Gln Tyr Cys Arg Tyr Gln Lys Cys Leu Ala Thr Gly Met Lys Arg
            260                 265                 270

Glu Ala Val Gln Glu Glu Arg Gln Arg Gly Lys Asp Lys Asp Gly Asp
        275                 280                 285

Gly Glu Gly Ala Gly Gly Ala Pro Glu Glu Met Pro Val Asp Arg Ile
    290                 295                 300

Leu Glu Ala Glu Leu Ala Val Glu Gln Lys Ser Asp Gln Gly Val Glu
305                 310                 315                 320

Gly Pro Gly Gly Thr Gly Gly Ser Gly Ser Ser Pro Asn Asp Pro Val
                325                 330                 335

Thr Asn Ile Cys Gln Ala Ala Asp Lys Gln Leu Phe Thr Leu Val Glu
            340                 345                 350

Trp Ala Lys Arg Ile Pro His Phe Ser Ser Leu Pro Leu Asp Asp Gln
        355                 360                 365
```

```
Val Ile Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe
    370                 375                 380

Ser His Arg Ser Ile Asp Val Arg Asp Gly Ile Leu Leu Ala Thr Gly
385                 390                 395                 400

Leu His Val His Arg Asn Ser Ala His Ser Ala Gly Val Gly Ala Ile
                405                 410                 415

Phe Asp Arg Val Leu Thr Glu Leu Val Ser Lys Met Arg Asp Met Arg
            420                 425                 430

Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ala Ile Ile Leu Phe Asn
        435                 440                 445

Pro Asp Ala Lys Gly Leu Ser Asn Pro Ser Glu Val Glu Val Leu Arg
    450                 455                 460

Glu Lys Val Tyr Ala Ser Leu Glu Thr Tyr Cys Lys Gln Lys Tyr Pro
465                 470                 475                 480

Glu Gln Gln Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu
                485                 490                 495

Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe Phe Lys Leu
            500                 505                 510

Ile Gly Asp Thr Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu Ala
        515                 520                 525

Pro His Gln Leu Ala
    530
```

<210> 25
<211> 1392
<212> DNA
<213> Homo sapiens

<400> 25

```
atgtatggaa attattctca cttcatgaag tttcccgcag gctatggagg ctcccctggc     60

cacactggct ctacatccat gagcccatca gcagccttgt ccacagggaa gccaatggac    120

agccacccca gctacacaga taccccagtg agtgccccac ggactctgag tgcagtgggg    180

acccccctca atgccctggg ctctccatat cgagtcatca cctctgccat gggcccaccc    240

tcaggagcac ttgcagcgcc tccaggaatc aacttggttg ccccacccag ctctcagcta    300

aatgtggtca acagtgtcag cagttcagag gacatcaagc ccttaccagg gcttcccggg    360

attggaaaca tgaactaccc atccaccagc cccggatctc tggttaaaca catctgtgct    420

atctgtggag acagatcctc aggaaagcac tacggggtat acagttgtga aggctgcaaa    480

gggttcttca agaggacgat aaggaaggac ctcatctaca cgtgtcggga taataaagac    540

tgcctcattg acaagcgtca gcgcaaccgc tgccagtact gtcgctatca gaagtgcctt    600

gtcatgggca tgaagaggga agctgtgcaa gaagaaagac agaggagccg agagcgagct    660

gagagtgagg cagaatgtgc taccagtggt catgaagaca tgcctgtgga gaggattcta    720
```

gaagctgaac ttgctgttga accaaagaca gaatcctatg gtgacatgaa tatggagaac 780

tcgacaaatg accctgttac caacatatgt catgctgctg acaagcagct tttcaccctc 840

gttgaatggg ccaagcgtat tccccacttc tctgacctca ccttggagga ccaggtcatt 900

ttgcttcggg cagggtggaa tgaattgctg attgcctctt tctcccaccg ctcagtttcc 960

gtgcaggatg gcatccttct ggccacgggt ttacatgtcc accggagcag tgcccacagt 1020

gctggggtcg gctccatctt tgacagagtt ctaactgagc tggtttccaa aatgaaagac 1080

atgcagatgg acaagtcgga actgggatgc ctgcgagcca ttgtactctt taacccagat 1140

gccaagggcc tgtccaaccc ctctgaggtg gagactctgc gagagaaggt ttatgccacc 1200

cttgaggcct acaccaagca gaagtatccg gaacagccag gcaggtttgc caagctgctg 1260

ctgcgcctcc cagctctgcg ttccattggc ttgaaatgcc tggagcacct cttcttcttc 1320

aagctcatcg gggacacccc cattgacacc ttcctcatgg agatgttgga gaccccgctg 1380

cagatcacct ga 1392

<210> 26
<211> 1392
<212> DNA
<213> Homo sapiens

<400> 26
tcaggtgatc tgcagcgggg tctccaacat ctccatgagg aaggtgtcaa tgggggtgtc 60

cccgatgagc ttgaagaaga agaggtgctc caggcatttc aagccaatgg aacgcagagc 120

tgggaggcgc agcagcagct tggcaaacct gcctggctgt tccggatact tctgcttggt 180

gtaggcctca agggtggcat aaaccttctc tcgcagagtc tccacctcag aggggttgga 240

caggcccttg gcatctgggt taaagagtac aatggctcgc aggcatccca gttccgactt 300

gtccatctgc atgtctttca ttttggaaac cagctcagtt agaactctgt caaagatgga 360

gccgacccca gcactgtggg cactgctccg gtggacatgt aaacccgtgg ccagaaggat 420

gccatcctgc acggaaactg agcggtggga gaaagaggca atcagcaatt cattccaccc 480

tgcccgaagc aaaatgacct ggtcctccaa ggtgaggtca gagaagtggg gaatacgctt 540

ggcccattca acgagggtga aaagctgctt gtcagcagca tgacatatgt tggtaacagg 600

gtcatttgtc gagttctcca tattcatgtc accataggat tctgtctttg gttcaacagc 660

aagttcagct tctagaatcc tctccacagg catgtcttca tgaccactgg tagcacattc 720

tgcctcactc tcagctcgct ctcggctcct ctgtctttct tcttgcacag cttccctctt 780

catgcccatg acaaggcact tctgatagcg acagtactgg cagcggttgc gctgacgctt 840

gtcaatgagg cagtctttat tatcccgaca cgtgtagatg aggtccttcc ttatcgtcct 900

cttgaagaac cctttgcagc cttcacaact gtataccccg tagtgctttc ctgaggatct 960

gtctccacag atagcacaga tgtgtttaac cagagatccg gggctggtgg atgggtagtt 1020

catgtttcca atcccgggaa gccctggtaa gggcttgatg tcctctgaac tgctgacact 1080

gttgaccaca tttagctgag agctgggtgg ggcaaccaag ttgattcctg gaggcgctgc 1140

aagtgctcct gagggtgggc ccatggcaga ggtgatgact cgatatggag agcccagggc 1200

attgaggggg gtccccactg cactcagagt ccgtggggca ctcactgggg tatctgtgta 1260

gctggggtgg ctgtccattg gcttccctgt ggacaaggct gctgatgggc tcatggatgt 1320

agagccagtg tggccagggg agcctccata gcctgcggga aacttcatga agtgagaata 1380

atttccatac at 1392

<210> 27
<211> 463
<212> PRT
<213> Homo sapiens

<400> 27

Met Tyr Gly Asn Tyr Ser His Phe Met Lys Phe Pro Ala Gly Tyr Gly
1 5 10 15

Gly Ser Pro Gly His Thr Gly Ser Thr Ser Met Ser Pro Ser Ala Ala
20 25 30

Leu Ser Thr Gly Lys Pro Met Asp Ser His Pro Ser Tyr Thr Asp Thr
35 40 45

Pro Val Ser Ala Pro Arg Thr Leu Ser Ala Val Gly Thr Pro Leu Asn
50 55 60

Ala Leu Gly Ser Pro Tyr Arg Val Ile Thr Ser Ala Met Gly Pro Pro
65 70 75 80

Ser Gly Ala Leu Ala Ala Pro Pro Gly Ile Asn Leu Val Ala Pro Pro
85 90 95

Ser Ser Gln Leu Asn Val Val Asn Ser Val Ser Ser Ser Glu Asp Ile
100 105 110

Lys Pro Leu Pro Gly Leu Pro Gly Ile Gly Asn Met Asn Tyr Pro Ser
115 120 125

Thr Ser Pro Gly Ser Leu Val Lys His Ile Cys Ala Ile Cys Gly Asp
130 135 140

Arg Ser Ser Gly Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys
145 150 155 160

Gly Phe Phe Lys Arg Thr Ile Arg Lys Asp Leu Ile Tyr Thr Cys Arg
165 170 175

Asp Asn Lys Asp Cys Leu Ile Asp Lys Arg Gln Arg Asn Arg Cys Gln
180 185 190

Tyr Cys Arg Tyr Gln Lys Cys Leu Val Met Gly Met Lys Arg Glu Ala
195 200 205

```
Val Gln Glu Glu Arg Gln Arg Ser Arg Glu Arg Ala Glu Ser Glu Ala
    210                 215                 220

Glu Cys Ala Thr Ser Gly His Glu Asp Met Pro Val Glu Arg Ile Leu
225                 230                 235                 240

Glu Ala Glu Leu Ala Val Glu Pro Lys Thr Glu Ser Tyr Gly Asp Met
                245                 250                 255

Asn Met Glu Asn Ser Thr Asn Asp Pro Val Thr Asn Ile Cys His Ala
            260                 265                 270

Ala Asp Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro
        275                 280                 285

His Phe Ser Asp Leu Thr Leu Glu Asp Gln Val Ile Leu Leu Arg Ala
    290                 295                 300

Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Val Ser
305                 310                 315                 320

Val Gln Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Ser
                325                 330                 335

Ser Ala His Ser Ala Gly Val Gly Ser Ile Phe Asp Arg Val Leu Thr
            340                 345                 350

Glu Leu Val Ser Lys Met Lys Asp Met Gln Met Asp Lys Ser Glu Leu
        355                 360                 365

Gly Cys Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ala Lys Gly Leu
    370                 375                 380

Ser Asn Pro Ser Glu Val Glu Thr Leu Arg Glu Lys Val Tyr Ala Thr
385                 390                 395                 400

Leu Glu Ala Tyr Thr Lys Gln Lys Tyr Pro Glu Gln Pro Gly Arg Phe
                405                 410                 415

Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys
            420                 425                 430

Cys Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile
        435                 440                 445

Asp Thr Phe Leu Met Glu Met Leu Glu Thr Pro Leu Gln Ile Thr
    450                 455                 460
```

<210> 28
<211> 1027
<212> DNA
<213> Homo sapiens

<400> 28

```
gactcccaag atggcggacc tactgggctc catcctgagc tccatggaga agccacccag     60

cctcggtgac caggagactc ggcgcaaggc ccgagaacag gccgcccgcc tgaagaaact    120

acaagagcaa gagaaacaac agaaagtgga gtttcgtaaa aggatggaga aggaggtgtc    180

agatttcatt caagacagtg ggcagatcaa gaaaaagttt cagccaatga acaagatcga    240

gaggagcata ctacatgatg tggtggaagt ggctggcctg acatccttct cctttqqqqa    300
```

agatgatgac tgtcgctatg tcatgatctt caaaaaggag tttgcaccct cagatgaaga 360

gctagactct taccgtcgtg gagaggaatg ggacccccag aaggctgagg agaagcggaa 420

gctgaaggag ctggcccaga ggcaagagga ggaggcagcc cagcaggggc ctgtggtggt 480

gagccctgcc agcgactaca aggacaagta cagccacctc atcggcaagg gagcagccaa 540

agacgcagcc cacatgctac aggccaataa gacctacggc tgtgtgcccg tggccaataa 600

gagggacaca cgctccattg aagaggctat gaatgagatc agagccaaga agcgtctgcg 660

gcagagtggg gaagagttgc cgccaacctc taggcgcccc gcccagctcc ctttgacccc 720

tggggcaggg caggggggcag ggagagacaa ggctgctgct attagagccc atcctggagc 780

cccacctctg aaccacctcc taccagctgt ccctcaggct gggggaaaac aggtgtttga 840

tttgtcaccg ttggagcttg gatatgtgcg tggcatgtgt gtgtgtgtgt gagagtgtga 900

atgcacaggt gggtatttaa tctgtattat tccccgttct tggaattttc ttccccatgg 960

ggctggggta ctttacattc aataaatact gtttaaccca aaaaaaaaaa aaaaaaaaaa 1020

aaaaaaa 1027

<210> 29
<211> 1027
<212> DNA
<213> Homo sapiens

<400> 29
ttttttttt ttttttttt ttttttttgg gttaaacagt atttattgaa tgtaaagtac 60

cccagcccca tggggaagaa aattccaaga acggggaata atacagatta aatacccacc 120

tgtgcattca cactctcaca cacacacaca catgccacgc acatatccaa gctccaacgg 180

tgacaaatca aacacctgtt ttcccccagc ctgagggaca gctggtagga ggtggttcag 240

aggtggggct ccaggatggg ctctaatagc agcagccttg tctctccctg ccccctgccc 300

tgccccaggg gtcaaaggga gctgggcggg gcgcctagag gttggcggca actcttcccc 360

actctgccgc agacgcttct tggctctgat ctcattcata gcctcttcaa tggagcgtgt 420

gtccctctta ttggccacgg gcacacagcc gtaggtctta ttggcctgta gcatgtgggc 480

tgcgtctttg gctgctccct tgccgatgag gtggctgtac ttgtccttgt agtcgctggc 540

agggctcacc accacaggcc cctgctgggc tgcctcctcc tcttgcctct gggccagctc 600

cttcagcttc cgcttctcct cagccttctg ggggtcccat tcctctccac gacggtaaga 660

gtctagctct tcatctgagg gtgcaaactc ctttttgaag atcatgacat agcgacagtc 720

atcatcttcc ccaaaggaga aggatgtcag gccagccact tccaccacat catgtagtat 780

gctcctctcg atcttgttca ttggctgaaa ctttttcttg atctgcccac tgtcttgaat 840

gaaatctgac acctccttct ccatcctttt acgaaactcc actttctgtt gtttctcttg 900

ctcttgtagt ttcttcaggc gggcggcctg ttctcgggcc ttgcgccgag tctcctggtc 960

accgaggctg ggtggcttct ccatggagct caggatggag cccagtaggt ccgccatctt 1020

gggagtc 1027

<210> 30
<211> 293
<212> PRT
<213> Homo sapiens

<400> 30

Met Ala Asp Leu Leu Gly Ser Ile Leu Ser Ser Met Glu Lys Pro Pro
1 5 10 15

Ser Leu Gly Asp Gln Glu Thr Arg Arg Lys Ala Arg Glu Gln Ala Ala
20 25 30

Arg Leu Lys Lys Leu Gln Glu Gln Glu Lys Gln Gln Lys Val Glu Phe
35 40 45

Arg Lys Arg Met Glu Lys Glu Val Ser Asp Phe Ile Gln Asp Ser Gly
50 55 60

Gln Ile Lys Lys Lys Phe Gln Pro Met Asn Lys Ile Glu Arg Ser Ile
65 70 75 80

Leu His Asp Val Val Glu Val Ala Gly Leu Thr Ser Phe Ser Phe Gly
85 90 95

Glu Asp Asp Asp Cys Arg Tyr Val Met Ile Phe Lys Lys Glu Phe Ala
100 105 110

Pro Ser Asp Glu Glu Leu Asp Ser Tyr Arg Arg Gly Glu Glu Trp Asp
115 120 125

Pro Gln Lys Ala Glu Glu Lys Arg Lys Leu Lys Glu Leu Ala Gln Arg
130 135 140

Gln Glu Glu Glu Ala Ala Gln Gln Gly Pro Val Val Val Ser Pro Ala
145 150 155 160

Ser Asp Tyr Lys Asp Lys Tyr Ser His Leu Ile Gly Lys Gly Ala Ala
165 170 175

Lys Asp Ala Ala His Met Leu Gln Ala Asn Lys Thr Tyr Gly Cys Val
180 185 190

Pro Val Ala Asn Lys Arg Asp Thr Arg Ser Ile Glu Glu Ala Met Asn
195 200 205

Glu Ile Arg Ala Lys Lys Arg Leu Arg Gln Ser Gly Glu Glu Leu Pro
210 215 220

Pro Thr Ser Arg Arg Pro Ala Gln Leu Pro Leu Thr Pro Gly Ala Gly
225 230 235 240

Gln Gly Ala Gly Arg Asp Lys Ala Ala Ala Ile Arg Ala His Pro Gly
245 250 255

```
Ala Pro Pro Leu Asn His Leu Leu Pro Ala Val Pro Gln Ala Gly Gly
            260                 265                 270

Lys Gln Val Phe Asp Leu Ser Pro Leu Glu Leu Gly Tyr Val Arg Gly
        275                 280                 285

Met Cys Val Cys Val
    290
```